



## Classification of PPMI MRI scans with voxel-based morphometry and machine learning to assist in the diagnosis of Parkinson's disease



Gabriel Solana-Lavalle, Roberto Rosas-Romero\*

Department of Electrical and Computer Engineering, Universidad de las Américas-Puebla Santa Catarina Mártir, San Andrés Cholula, Puebla, 72810, México

### ARTICLE INFO

#### Article history:

Received 21 August 2020

Accepted 6 October 2020

#### Keywords:

Parkinson's disease  
Magnetic resonance imaging  
Voxel-based morphometry  
Machine learning  
Feature selection

### ABSTRACT

**Background and objectives:** Qualitative and quantitative analyses of Magnetic Resonance Imaging (MRI) scans are carried out to study and understand *Parkinson's Disease*, the second most common neurodegenerative disorder in people at their 60's. Some quantitative analyses are based on the application of *voxel-based morphometry* (VBM) on *magnetic resonance* images to determine the *regions of interest*, within gray matter, where there is a loss of the nerve cells that generate dopamine. This loss of dopamine is indicative of *Parkinson's disease*. The purpose of this research is the introduction of a new method to classify the 3-D magnetic resonance scans of an individual, as an assisting tool for diagnosis of *Parkinson's disease* by using the largest MRI dataset (*Parkinson's Progression Markers Initiative*) from a population of patients with *Parkinson's disease* and control individuals. A contribution is that separate studies are conducted for men and women since gender plays a significant role within Neurobiology, which is demonstrated by the fact that men are more prone to *Parkinson's disease* than women are.

**Methods:** Previous to classification, VBM is conducted on magnetic resonance images to detect the regions where features are extracted by using first- and second-order statistics methods. Furthermore, the number of features is considerably reduced by using *feature selection techniques*. Seven classifiers are used and we are conducting separate experiments for men and women.

**Results:** The best detection performance achieved in men is 99.01% of accuracy, 99.35% of sensitivity, 100% of specificity, and 100% of precision. The best detection performance achieved in women is 96.97% of accuracy, 100% of sensitivity, 96.15% of specificity, and 97.22% of precision. During classification of magnetic resonance images, the corresponding computational complexity is reduced since few features are selected.

**Conclusions:** The proposed method provides high performance as an assisting tool in the diagnosis of *Parkinson's disease*, by conducting separate experiments in men and women. While previous works have focused their analysis to the striatum region of the brain (the largest nuclear complex of the basal ganglia), the proposed approach is based on analysis over the whole brain by looking for decreases of tissue thickness, with the consequence of finding other regions of interest such as the cortex.

© 2020 The Author(s). Published by Elsevier B.V.  
This is an open access article under the CC BY-NC-ND license  
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## 1. Introduction

*Parkinson's disease* (PD) is the second most common neurodegenerative disorder for people at their 60's. The characteristic symptoms are tremor at rest, bradykinesia, and rigidity. The cause of PD is the loss of dopaminergic neurons in the substantia nigra of the brain [1].

*Magnetic Resonance Imaging* (MRI) of the brain has been used to assist the diagnosis of *Parkinson's disease* (PD) with an increase in the detection accuracy over the last years [2,3]. The brain structure, over different types of tissue, has been visually and qualitatively interpreted by physicians on MR images; however, an automated quantitative analysis is also recommended since it improves the interpretation of the changes in brain structure [2–6].

To conduct this research, we used MRI images from the *Parkinson's Progression Markers Initiative* (PPMI), sponsored by the *Michael J. Fox Foundation*. One result, of this initiative, is one of the most

\* Corresponding author.

E-mail address: [roberto.rosas@udlap.mx](mailto:roberto.rosas@udlap.mx) (R. Rosas-Romero).

robust libraries of clinical and imaging data and biosamples for Parkinson's disease research [1].

*Voxel-Based morphometry* (VBM) is an approach to compute the differences, within regional concentrations of gray matter, through a voxel-wise comparison of 3-D brain images between two groups of subjects [7,8]. To conduct VBM, images, from multiple subjects, are normalized (contrast stretched) and registered to generate a brain template or atlas that represents a particular group of subjects. The gray matter, for the template, is segmented and smoothed. VBM has been used to investigate the differences in brain structure between (1) a group of PD patients and a group of healthy individuals or controls; and (2) the same group of PD patients at different time instances [3,9]. The findings of the application of VBM are that PD patients are characterized by a decrease in gray matter volume, when compared with controls [3,9,10]. A decrease of gray matter volume has also been observed within the same group of PD patients after some periods of time [9]. Another finding has to do with the brain regions where a decrease of gray matter volume takes place, being putamen and parietal cortex those regions where significant differences are identified [3]. Finally, biomarkers (clinical signs) for PD can be determined by applying VBM on MRI studies [11]. Validated PD biomarkers would accelerate PD therapeutics research.

One motivation for this work is to identify and classify gray matter regions, which are affected by the disorder through conducting separate studies, for men and women. It has been found that Parkinson's disease is more likely in men than in women because of neurobiologic differences [12–14]. For instance, some female predominant neuro-transmitters and hormones, such as estrogen and oxytocin, end up being a protection for the female nervous system. On the other hand, the small number of subjects in brain MRI datasets has been a deterrent to investigate PD detection on separate groups, men and women; nonetheless, the PPMI MRI dataset is one of the most complete and appropriate datasets for such task.

Another motivation is the application of machine learning and the automated classification of PPMI MRI data since there have been few efforts for classification of MR images for PD detection. There have been successful studies where Non-MRI images were automatically analyzed to assess PD disease progression [15–17]. Classification of PD patients and control subjects has been done by using a Support Vector Machine (SVM) fed with features extracted from resting-state functional MRI (rsfMRI) images [18]; however, the used dataset is considerably smaller (nineteen PD patients and twenty-seven healthy subjects) than the PPMI MRI dataset, experiments were not conducted in separate populations (men and women), and their performance is lower than the one from the proposed method (accuracy of 86.96%, sensitivity of 78.95%, specificity of 92.59%). A sparse feature selection framework has been proposed along with a multi-class classification model by using multi-modality data from the PPMI dataset with an accuracy of 78.37%  $\pm$  8.11% [19]. A Convolutional Neural Network (CNN) has been used to refine the diagnosis of Parkinson's disease [20], where the MR images are tested to give an accuracy of 88.9% by using the PPMI dataset. A deep learning framework was proposed for simultaneous classification and regression of Parkinson disease diagnosis based on PPMI MR images and personal information (age, gender) with an accuracy of 100% [21], and without carrying out separate experiments for men and women. Another CNN was used to classify PD and healthy controls from the PPMI MRI dataset with a better performance from 3%–9% in terms of accuracy, sensitivity, and specificity when compared to the other techniques [22]. MRI images from 28 PD patients, 28 progressive supranuclear palsy patients and 28 healthy control subjects were used by an algorithm based on the combination of Principal Components Analysis, for feature extraction, and on SVM, for classifi-

cation, with an accuracy, specificity and sensitivity above 90% for the case of PD patients vs. controls [23]. Another strategy consists in using a CNN to classify PD from Neuromelanine sensitive MRI while achieving 85% of accuracy [24]. Neuromelanine sensitive MRI is an alternative to structural MRI, as neuromelanin is a neuronal pigment. Another SVM was used to classify PD patients in a prodromal phase (from the PPMI data set) by using network features (NFs) and clinical features (CFs) while achieving 93% of accuracy, 93% of sensitivity, and 92% of specificity [25].

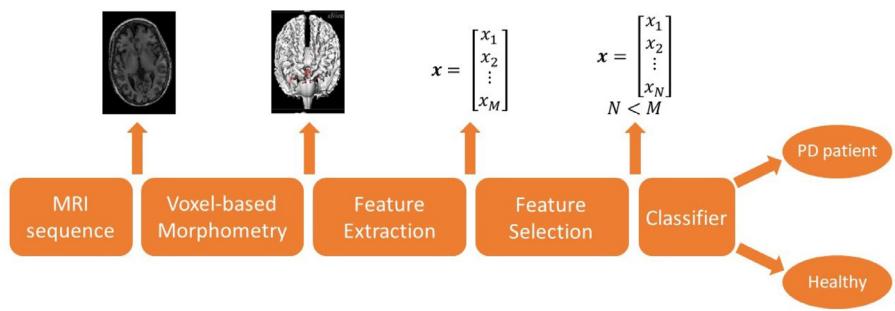
Voice-based analysis and machine learning have also been applied as an automatic decision tool to help physicians interpret functional auditory information and decide whether there is PD or not [26–30]. Since PD affects the ability of the patients to move, PD has also been predicted by extracting features from handwritten exams by using image processing techniques and an improved and optimized version of the crow search algorithm (OCSA) with an accuracy of 100% for 20 benchmark datasets [31]. A review concerning PD detection by means of recent technologies was presented by Pereira et al. [32], where several approaches based on image and signal analysis, smartphone devices, virtual and augmented reality, sensors and web-based application were considered, along with datasets widely used.

Gender plays a significant role for the optimal discrimination between PD patients and controls [33], and subsequent interpretation of results. In this work, classification of 3D MRI scans for PD detection is developed and applied to each gender as well as the subsequent interpretation. This is possible since we are using the largest dataset available, where the number of observations is adequate for partitioning of the dataset according to gender. The use of separate sets, for male and female subjects, allows to find the most relevant regions of interest for each gender. Another contribution, of our work, consists in exploring (1) the feasibility of second-order statistics for feature extraction, (2) the use of feature selection techniques to find the most relevant features while reducing computational complexity, and (3) the application of multiple classifiers (*Logistic, Naive Bayes, k Nearest Neighbors, Bayesian Network, Random Forest, Multi-Layer Perceptron, Support Vector Machine*) for a binary decision. The binary output, from a classifier, is of clinical utility when the regions, under analysis, are also specified since clinicians need to understand the basis for a recommendation by visually examining those regions on images. The proposed method is characterized by a high classification performance. The major differences from previous works are (1) the conduction of separate gender-based studies, (2) the use of multiple classifiers to measure PD detection performance, (3) the achievement of high detection accuracy while using a small number of features (low complexity), (4) the extraction of ROIs, for further analysis, in brain locations different from those usually used for PD detection.

The rest of the paper is organized as follows. Section 2 describes the methodology used while conducting this research. The results obtained are reported in Section 3 and discussed in Section 4. Conclusions are presented in Section 5.

## 2. Methodology

The proposed approach for Parkinson's disease detection on Magnetic Resonance Imaging (MRI) consists of four main stages: (1) *detection of regions of interest*, (2) *feature extraction*, (3) *feature selection*, (4) *classification*, and (5) *performance assessment*. Fig. 1 shows the first four main stages during PD detection. Regions of interest have to be determined for feature extraction within gray matter. Regions of interest are those areas where a decrease of volume in gray matter is identified. VBM is used to obtain these regions of interest by comparing two templates representing two groups, healthy individuals and PD patients. In general, the pro-

**Fig. 1.** Stages for voxel-based morphometry PD detection.

posed method consists of two main parts: (1) Identification of regions of interest by using VBM and (2) analysis of these regions for PD detection. After VBM, the pattern recognition pipeline consists of (1) *feature extraction* based on first- and second-order statistics, (2) *feature selection* based on *Principal Component Analysis* (PCA) followed by *wrapper feature selection*, and (3) classification with *10-fold cross validation* based on seven different classifiers (*K*-nearest neighbors with *K* = 5, multi-layer perceptron, support vector machine with a radial basis kernel, random forests, naïve Bayes classifier, logistic classifier and Bayesian networks). The following subsections explain the different stages.

## 2.1. Dataset

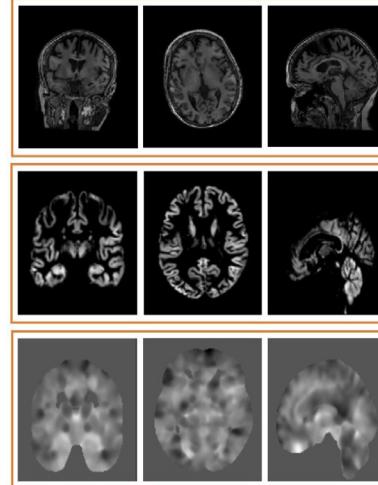
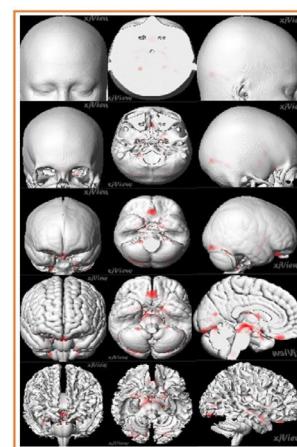
To conduct experiments, images, from patients and controls, were accessed from the *Parkinson's Progression Makers Initiative* (PPMI) dataset. The PPMI is being conducted at a network of clinical sites around the world. One goal of the PPMI is to collect clinical, genomics, patient-reported and imaging data, which is publicly shared to accelerate research discoveries in the definition of biomarkers of PD progression. These biomarkers can be used in therapeutic studies, which is the ultimate goal [1].

PPMI's T1-weighted MR images were selected for this study. These images were generated by using a 1.5–3 Tesla scanner. The total scan time is in the range between 20–30 min. The T1-weighted MR images were acquired as a 3D sequence with a slice thickness of 1.5 mm or less, under three different views: axial, sagittal and coronal. The MR images were extracted from 226 male patients with PD, 86 male controls, 104 female patients with PD, and 64 female patients with PD.

## 2.2. Voxel-based morphometry for detection of regions of interest

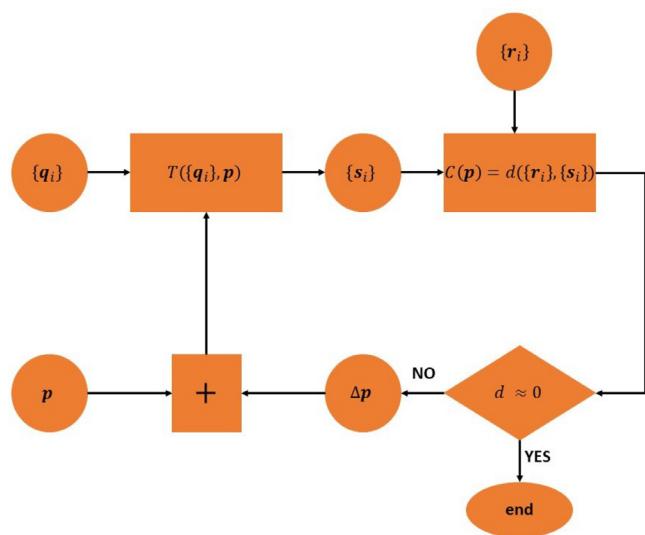
*Voxel-based morphometry* (VBM), applied to MR images, is used for extraction of regions of interest within gray matter. During VBM of MR images, five steps are taken: (1) For each subject within a group, the MRI planes (axial, sagittal, coronal) are registered or aligned. (2) The gray matter, from aligned images, is then segmented, while white matter and cerebrospinal fluid are also segmented, but they are discarded. (3) The aligned and segmented gray matter is smoothed. (4) For each group, a template is generated by averaging the aligned, segmented and smoothed MR images from all the subjects of the same group. (5) A voxel-wise analysis is conducted to determine and localize differences (regions of interest) between the two groups of subjects, where each group is represented by a template or atlas.

The result of VBM is a map (known as *t-map*) that shows the voxel differences of gray matter concentrations between the templates representing two groups of subjects. Fig. 2 shows three MRI planes (coronal, axial and sagittal) in the first row along with the result of alignment-segmentation-smoothing (second row), and the

**Fig. 2.** Three MRI planes: coronal (left column), axial (middle), sagittal (right column). Original Images (top row). Result after normalization and segmentation (second row). *T*-map that shows the result of comparing two templates (bottom row).**Fig. 3.** 3-D reconstruction from MRI (for each plane or view) with the regions of interest detected by using a two-sample *t*-test. Each column corresponds to one view: coronal (left), axial (middle), sagittal (right).

*t*-map (third row) that is the result of computing the differences between the templates obtained from PD patients and controls. The *t*-map shows values, which are used to determine regions of interest. Voxels, in the *t*-map, are labeled as part of regions of interest according to a two-sample *t*-test threshold.

A 3-D reconstruction of one subject is shown in Fig. 3, which includes cross-sectional views over different planes (coronal, axial and sagittal). The highlighted regions, in red color, correspond to



**Fig. 4.** Block diagram for the optimization process to align a set of voxels to a reference set of voxels.

regions where there is a significant decrease in gray matter concentration.

### 2.2.1. Image alignment

Deformation is the alteration of shapes through a *transformation* of the location of voxels (3-D points with gray level intensity) belonging to an object according to  $\mathbf{s}_i = T(\mathbf{q}_i, \mathbf{p})$ , where a voxel position  $\mathbf{q}_i$  is assigned a new value  $\mathbf{s}_i$  depending on the set of transformation parameters  $\mathbf{p}$ . For the case of an *affine transformation*,  $\mathbf{p}$  corresponds to angles for object *rotation*, scalars for object *scaling* (increasing or decreasing) and offsets for *spatial shift* of the object. An affine transformation takes the form

$$\begin{bmatrix} r_x \\ r_y \\ r_z \end{bmatrix} = \begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} \\ p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} \\ p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} \end{bmatrix} \begin{bmatrix} q_x \\ q_y \\ q_z \\ 1 \end{bmatrix} \quad (1)$$

Aligning images consists of estimating the transformation parameters  $\mathbf{p}$  that optimize the match of a set of transformed voxels  $\{\mathbf{s}_i\}$  to another set of reference voxels  $\{\mathbf{r}_i\}$ , where both sets are extracted from an observed object  $O_1$  under transformation  $T(O_1, \mathbf{p})$  and a reference object  $O_2$ , respectively. During alignment, affine transformations are repeatedly applied to  $O_1$ , by adjusting the set of parameters  $\mathbf{p}$ , until it becomes the most similar to  $O_2$ , which is the same as minimizing the *cost function*, defined by using the *Euclidean distance* between  $T(O_1, \mathbf{p})$  and  $O_2$ , according to  $C(\mathbf{p}) = \sum_{i=1}^N \|T(\mathbf{q}_i, \mathbf{p}) - \mathbf{r}_i\|^2$ , where  $N$  is the number of object voxels. A general block diagram to find the optimal alignment between two sets of voxels is given in Fig. 4. The cost function  $C(\mathbf{p})$  can be iteratively minimized by using the *gradient descent* method or the *Levenberg-Marquardt* method for computation of  $\Delta\mathbf{p}$  in  $\mathbf{p}^{(k+1)} = \mathbf{p}^{(k)} + \Delta\mathbf{p}$ .

The result of aligning a MR image is shown in Fig. 5. Since an object consists of a discrete set of voxels located in a 3-D grid, the application of affine transformations to voxels implies the change of voxel positions. Thus, the use of a regression model is required to interpolate voxel intensity values into the original 3-D grid.

### 2.2.2. MR image segmentation

Three-dimensional MR image segmentation consists in the classification of voxels according to the tissue (white matter, gray matter, cerebrospinal fluid) each voxel is more likely to belong

to. A 3-D tissue probability map is learned to assist the classification of voxels by modeling a cost function as a *mixture of K Gaussian functions*. The number of classes  $K$  corresponds to the number of tissue types. The probability of a voxel with intensity  $x_i$ , given that it belongs to tissue  $k$ , is given by  $P(x_i|k, \mu_k, \sigma_k) =$

$\frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(x_i-\mu_k)^2}{2\sigma_k^2}}$ , where the Gaussian function, for tissue  $k$ , is characterized by its *average value*  $\mu_k$  and *variance*  $\sigma_k$ . The Gaussian functions are mixed through a set of weights  $\{\gamma_k\}$ , where  $\sum_{k=1}^K \gamma_k = 1$ . The prior probability of any voxel belonging to the  $k$ th tissue is given by the proportion of voxels,  $P(k|\gamma_k) = \gamma_k$ . The joint probability of a voxel with intensity  $x_i$  in cluster  $k$  is given by  $P(x_i, k|\mu_k, \sigma_k, \gamma_k) = P(x_i|k, \mu_k, \sigma_k)P(k|\gamma_k)$ . A voxel of intensity  $x_i$  is assigned to the tissue with the highest joint probability. The probability of voxel value  $x_i$ , given the set of parameters, is obtained by summing over all the  $K$  tissue clusters,  $P(x_i|\mu, \sigma, \gamma) = \sum_{k=1}^K P(x_i, k|\mu_k, \sigma_k, \gamma_k)$ . The learning of the parameters of the *mixture of Gaussians* (MOG) model  $\{\mu_k, \sigma_k, \gamma_k | k = 1, 2, \dots, K\}$  consists in minimizing the probability of observing all the voxel values  $\{x_i; i = 1, 2, \dots, I\}$  given the parameters of the  $K$  Gaussians, according to

$$P(x_1, \dots, x_I | \mu, \sigma, \gamma) = \prod_{i=1}^I P(x_i | \mu, \sigma, \gamma) P(k|\gamma_k) = \prod_{i=1}^I \left[ \sum_{k=1}^K \frac{\gamma_k}{\sqrt{2\pi}\sigma_k} e^{-\frac{(x_i-\mu_k)^2}{2\sigma_k^2}} \right]. \quad (2)$$

The logarithm of the previous cost function results in a monotonically related function to be maximized

$$\xi = - \sum_{i=1}^I \log \left[ \sum_{k=1}^K \frac{\gamma_k}{\sqrt{2\pi}\sigma_k} \exp^{-\frac{(x_i-\mu_k)^2}{2\sigma_k^2}} \right]. \quad (3)$$

The model parameters are obtained by solving  $\arg \max_{\mathbf{p}} \xi$ . The result of segmenting a MR image, along with the alignment outcome, are shown in Fig. 5.

### 2.2.3. Extraction of regions of interest

To compute the voxel-wise statistical differences between the two groups under comparison (PD patients and controls), a *two-sample t-test* was used. For this statistical analysis, it is assumed that the voxel intensity values, from each group, are independent and normally distributed. During the *two-sample t-test*, the difference between the two population means is tested to determine whether the means are equal or their difference is significant. The default *null hypothesis*  $H_0$  for the two-sample *t-test* is that the two population means are equal ( $\mu_1 = \mu_2$ ). Hence, to test whether the means are different  $H_1$  ( $\mu_1 \neq \mu_2$ ), the *t-statistic* value is computed according to

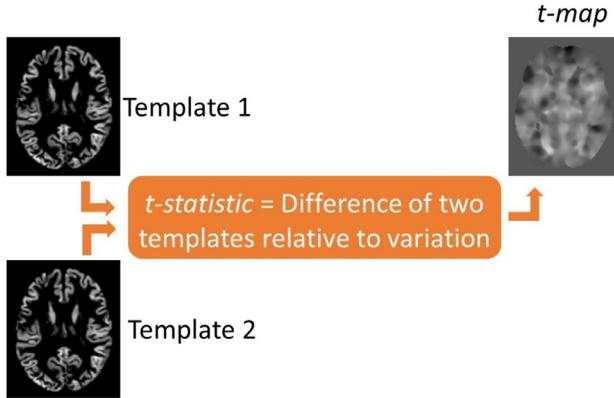
$$t = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} \quad (4)$$

where  $N_1$  and  $N_2$  are the number of subjects within each group, PD patients and controls. The result of the voxel-wise computation of the *t*-statistic, from two templates, is shown in Fig. 6.

The *p-value* is used to determine the regions of interest, where differences of gray matter concentrations are significant between the two groups, such as the case of the red regions in Fig. 3. The *p-value* is the area under a probability curve at the tail of the curve, starting at the *t*-statistic. The *p-value* is related to the *t*-statistic in such a way that the larger the absolute value of the *t*-statistic, the smaller the *p-value*, and the greater the evidence against the null hypothesis. The *p-value* for this work was selected as 0.001. Besides the *p-value*, an additional requirement to determine regions of interest is the number of connected voxels. The motivation for the use of connectivity is to avoid the selection of noisy regions.



**Fig. 5.** An original MR image and the corresponding result after alignment and segmentation of the gray matter.



**Fig. 6.** Generation of a *t-map* from the templates of two groups, PD patients and controls.

### 2.3. Feature extraction from regions of interest

After the definition of 3-D regions of interest for analysis, features can be extracted from any subject at positions within the already established regions so that a decision can be taken to assist PD detection. Useful information, for classification over these regions, is defined by the way pixel intensity values are distributed. We are referring to first-order and second-order statistics of the regions under analysis.

The *first-order histogram* of a region of interest is given by

$$P(x_i) = \frac{\text{number of voxels with intensity } x_i}{\text{number of voxels within the regions of interest}} \quad (5)$$

where the first-order histogram  $\{P(x_i); i = 1, 2, \dots, N_v\}$  is a vector and the number of entries equals the number of pixel intensity values  $N_v$ . Multiple features are based on the first-order histogram, such as the *moments*  $\mu_k = E[x_i^k] = \sum_{i=1}^{N_v} x_i^k P(x_i)$ , the *central moments*  $c_k = E[(x_i - \mu_1)^k] = \sum_{i=1}^{N_v} (x_i - \mu_1)^k P(x_i)$ , and *entropy*  $H = E[-\log P(x_i)] = -\sum_{i=1}^{N_v} P(x_i) \log P(x_i)$ .

The *second-order statistics* features carry information about the relative positions of the pixel intensity values within the regions of interest. The computation of these features is based on the *gray level co-occurrence matrix* (GLCM) or *Haralick texture method*, where pairs of pixel values are analyzed in terms of the distance between them and the angle of the line that joins both pixels. The angles for generation of the GLCM are  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  and  $135^\circ$ , which correspond to orientation in four directions: horizontal, diagonal, vertical and antidiagonal. The GLCM for a pair of two voxel intensity values  $(x_i, x_j)$  at distance  $d$  and orientation  $\phi$  is defined as

$$P(x_i, x_j, d, \phi) = \frac{\text{Number of pairs } (x_i, x_j) \text{ at distance } d \text{ and angle } \phi}{\text{Total number of possible pairs}} \quad (6)$$

where  $\{P(x_i, x_j); i = 1, 2, \dots, N_v; j = 1, 2, \dots, N_v\}$ , at distance  $d$  and angle  $\phi$ , is also called *co-occurrence matrix*, which is the probability of occurrence of two pixel intensity values in terms of their relative position. The GLCM is generated for each region of interest at each view (axial, sagittal and coronal). There are 28 gray-level invariant Haralick texture features obtained from the GLCM [34,35]. Some of the most important measures, arising from the second-order histogram, are the *angular second moment*  $ASM = \sum_{i=1}^{N_v} \sum_{j=1}^{N_v} P(x_i, x_j)^2$ , *contrast*  $CON = \sum_{n=1}^{N_v} n^2 \left[ \sum_{i=1}^{N_v} \sum_{j=1, |x_i-x_j|<n}^{N_v} P(x_i, x_j) \right]$ , *inverse different moment*  $IDF = \sum_{i=1}^{N_v} \sum_{j=1}^{N_v} \frac{P(x_i, x_j)}{1+(x_i-x_j)^2}$ , *entropy*  $H = -\sum_{i=1}^{N_v} \sum_{j=1}^{N_v} P(x_i, x_j) \log P(x_i, x_j)$ , and *correlation*  $Corr = \frac{\sum_i \sum_j x_i x_j P(x_i, x_j) - \mu_i \mu_j}{\sigma_i \sigma_j}$ .

### 2.4. Feature selection

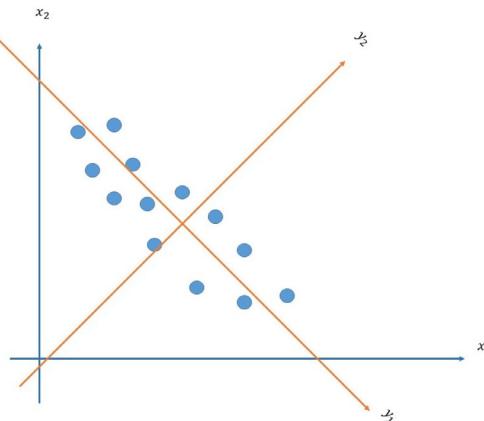
The number of features extracted from one subject's template amounts to thousands of features. The reason for this has to do with the number of regions of interest (more than ten regions of interest for male and more than twenty in women), the number of directions for second-order statistics features (0°, 45°, 90°, 135°), the number of views (sagittal, axial, coronal), number of different first-order statistics features, and number of second-order statistics features. It is reasonable to conclude that many of these features do not contribute with relevant information to assist in the detection of PD, are noisy and/or correlated. Two selection strategies were used, *Principal Component Analysis* followed by *Wrappers feature subset selection*, to: (1) reduce the number of features fed to classifiers, (2) find the most relevant and uncorrelated features, and (3) reduce computational complexity,

#### 2.4.1. Principal component analysis

We searched for some of the most important features by using *Principal Component Analysis* (PCA), a process to build a new feature space basis for a set of observations where the amount of variation in the observations over the principal components is maximized. Thus, PCA is a transformation where each feature vector  $x$  is transformed into feature vector  $y$  with a representation on the new feature space, as it is shown in Fig. 7 for the case of two-dimensional feature vectors.

Before applying PCA to the original features  $\{x_1, x_2, \dots, x_p\}$ , the average value of the population of feature vectors is subtracted from each vector. The  $i$ th component  $y_i$ , which results of applying PCA, is a linear combination of the original features  $y_i = \phi_{i,1} x_1 + \phi_{i,2} x_2 + \dots + \phi_{i,p} x_p$  that gives the largest possible variance  $\bar{y}_i^2$ . The weights  $\{\phi_{i,1}, \phi_{i,2}, \dots, \phi_{i,p}\}$  are called *loadings* and the new features  $\{y_1, y_2, \dots, y_p\}$  are called *scores*.

The weight vectors  $\phi_1, \phi_2, \dots, \phi_p$  are obtained by computing the *eigenvectors*  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p\}$  of the covariance matrix  $\mathbf{X}\mathbf{X}^T$ ; where the columns of matrix  $\mathbf{X}$  are the feature vectors before PCA. The *eigenvalues*  $\{\lambda_1, \lambda_2, \dots, \lambda_p\}$  determine the significance of each



**Fig. 7.** Set of 2-D feature vectors before ( $x_1$  vs.  $x_2$ ) and after ( $y_1$  vs.  $y_2$ ) PCA.

score  $\{y_1, y_2, \dots, y_p\}$ . The significance for score  $y_i$ , known as *percentage of the score variance*, is computed as  $\frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_p}$ . The loading component  $\phi_{ij}$  defines the percentage of the contribution of feature  $x_j$  to each score  $y_i$ . Those features  $\{x_j\}$ , with the highest contribution for PD detection, are found in two steps: (1) The most principal scores  $\{y_i\}$  are identified as those where the sum of their variances  $\{\frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_p}\}$  is higher than 90%. The selected scores are identified by a set of indexes,  $\{k\}$ . (2) Finally, the selected features  $\{x_j\}$  are those that achieve the highest value according to  $\sum_k \frac{\lambda_k \phi_{k,j}}{\lambda_1 + \lambda_2 + \dots + \lambda_p}$ .

#### 2.4.2. Feature subset selection

The number of selected feature, after PCA, still remains very high, above one thousand, and this is why there is a need for further reduction. At this point, the application of *feature subset selection* (search algorithms based on combinatorial analysis of very high computational complexity) is more feasible.

The criterion for selecting the best reduced subset of features is based on the subset that provides the best classification performance. During *feature subset selection*, different classifiers are trained by feeding them with each possible subset of  $k$  features, where  $k = 1, 2, \dots, p$  and  $p$  is the total number of features. First, a separate classifier is fed with one feature ( $k = 1$ ), where the total number of one-feature cases is  $p$ , and the classifier with the best performance is selected. Later, different classifiers are fed with two features ( $\frac{p(p-1)}{2}$  cases) and the best performing two-feature classifier is selected. The process is repeated until all the  $p$  features are applied to one classifier and its performance is measured. Finally, the selected classifiers (at each  $k$  value) are analyzed to determine the best feature subset. The problem of this selecting process is that the number of cases, to be analyzed, is  $2^p - 1$ . Hence, heuristics are introduced to avoid analyzing all the combinations of features. *Stepwise selection* techniques are based on heuristics to quickly select a feature subset by analyzing a reduced number of feasible subsets.

*Forward stepwise selection* begins with an empty set of features. At each iteration, one feature is added to the set, according to its relevance. At the final iteration, all the features are included. At each iteration, the feature, to be added, is the one that belongs to the subset with the best classification performance. The final subset is the best among the subsets from all iterations in terms of classification performance. Another strategy, called *backward stepwise selection*, starts with the complete set of features, and then the least relevant feature is removed, one at each iteration. Instead of analyzing  $2^p - 1$  feature subsets, stepwise selection tech-

niques analyze a much smaller number of subsets, which is equal to  $1 + \frac{p(p+1)}{2}$ .

*Wrappers feature subset selection* [36] was used in this work. This strategy combines the forward and backward stepwise selection techniques. At each step, a new feature is added while the relevance of already added features is also checked. When a previously included feature is determined as insignificant, it is removed.

#### 2.5. Classifiers

Different classification techniques were used in this work, the  $K$ -nearest neighbors, multi-layer perceptron, support vector machine, random forest, naive Bayes classifier, logistic classifier, and Bayesian networks.

##### 2.5.1. *K*-nearest neighbors

The *K-Nearest Neighbor* (KNN) is a non-linear classifier. The learning of a KNN consists of storing each training observation along with its corresponding class  $\{x_i, \ell_i; i = 1, 2, \dots, N\}$ . To assign a class to an unknown feature vector  $\mathbf{x}$ , the Euclidean distance between  $\mathbf{x}$  and each stored feature vector, is computed according to  $D_i = \sqrt{(\mathbf{x} - \mathbf{x}_i)^T (\mathbf{x} - \mathbf{x}_i)}$ . The  $K$  smallest distance values correspond to the  $K$  nearest neighbors to  $\mathbf{x}$ . The class that is assigned to  $\mathbf{x}$  is the most occurring class among the  $K$  nearest neighbors, where  $K$  must be an odd integer number to avoid a tie. In this research,  $K = 5$ .

##### 2.5.2. Multi-layer perceptron

The *Multi-Layer Perceptron* (MLP) is based on the perceptron algorithm by Rosenblatt [37]. The *perceptron* is the basic neural unit within a MLP. The perceptron consists of a linear combination of the incoming features  $x_1, x_2, \dots, x_p$ , followed by an *activation function*, according to  $y = f(\omega_0 + \sum_{i=1}^p \omega_i x_i)$ ; where the parameters  $\omega_0, \omega_1, \dots, \omega_p$  are called *synaptic weights* and the activation function is defined as the *logistic function*  $f(v) = \frac{1}{1+e^{-\alpha v}}$ . The MLP consists of multiple layers of neurons. At each layer, a neuron is fully connected to all the neurons in the next layer. The *back propagation* algorithm is used to train a MLP, which consists in adjusting the synaptic weights by using the *gradient descent* method at each iteration, according to  $\omega_j(n+1) = \omega_j(n) - \eta \frac{\partial C}{\partial \omega_j}$ ; where  $C$  is the *cost function* to be minimized and  $\eta$  is the *learning rate* that affects how fast the back-propagation will reach a minimum in the cost function.

##### 2.5.3. Support vector machine

The *Support Vector Machine* (SVM) is a supervised learning machine where the optimal hyperplane, among classes, is found. The SVM model is defined as  $y = b + \sum_{i=1}^N \omega_i K(\mathbf{x}, \mathbf{x}_i)$ ; where  $\mathbf{x}$  is the feature vector under classification,  $N$  is the total number of feature vectors,  $\mathbf{x}_i$  is the  $i$ th *support vector*,  $\omega_1, \dots, \omega_p$  are the parameters to be learned along with the *bias*  $b$ , and  $K(\mathbf{x}, \mathbf{x}_i)$  is a *kernel function*. The chosen kernel was the *Radial Basis function*,  $K(\mathbf{x}, \mathbf{x}_i) = e^{-\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{x}_i\|^2}$  ( $\sigma = 0.1$ ). The architecture of the SVM consists of (1) one output (since there are two classes, PD and healthy), and (2) 8-20 input nodes. The *sequential minimal optimization* (SMO) algorithm is used to train the SVM. The convergence of the SMO is tested by checking whether the *Karush-Kuhn-Tucker* (KKT) conditions are satisfied to within a tolerance value set to 0.1.

##### 2.5.4. Random forest

The *Random Forest* (RF) is a supervised learning machine that consists of a large number of *decision trees*. A decision tree partitions the feature space into multiple regions, where each region corresponds to a different class or leaf. A decision tree is characterized by a high variance. Each decision tree splits the feature space

quite different from others, since each decision tree is trained in a different way, by randomly picking feature vectors (*bagging*). RF consists of multiple decision trees, where each decision tree classifies the feature vector and the class assigned is the most voted by the decision trees. For this research, the RF implementation consisted of a set of 87–120 decision trees by reaching a tree depth of 150.

#### 2.5.5. Naive bayes

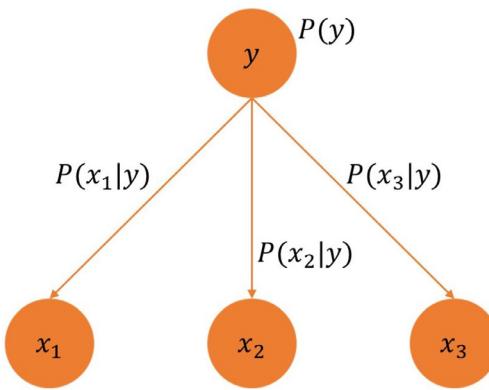
The *Naive Bayes* classifier model is based on a set of functions. These functions assign to each class  $C_k$  a probability value given a feature vector, according to  $P(C_k|x_1, x_2, \dots, x_p)$ , under the assumption that these features are independent. This conditional probability is reformulated by using the Bayes' theorem  $P(C_k|x_1, x_2, \dots, x_p) = \frac{P(x_1, x_2, \dots, x_p|C_k)P(C_k)}{P(x_1, x_2, \dots, x_p)}$ , where  $P(C_k)$  is called *prior*,  $P(x_1, x_2, \dots, x_p|C_k)$  is the *likelihood* and  $P(C_k|x_1, x_2, \dots, x_p)$  is the *posterior*. In the Naive Bayes classifier, the class assigned to a feature vector corresponds to the maximum posterior, according to  $\text{class} = \arg \max_k P(C_k|x_1, x_2, \dots, x_p)$ , which is a decision rule called *maximum a posteriori*. The denominator  $P(x_1, x_2, \dots, x_p)$  is common to all functions, and that is why its computation is not necessary during classification. Under the assumption that the features are mutually independent, the maximum posterior is found by computing  $P(C_k|\mathbf{x}) \propto P(C_k) \prod_{i=1}^p P(x_i|C_k)$ . The prior is computed as the proportion of observations belonging to each class. The conditional probability  $P(x_i|C_k)$  is modeled as a Gaussian,  $P(x_i|C_k) = \frac{1}{\sqrt{2\pi}\sigma_{k,i}} e^{-\frac{(x_i-\mu_{k,i})^2}{2\sigma_{k,i}^2}}$ , where  $\mu_{k,i}$  is the mean of the values in feature  $x_i$  associated to class  $C_k$ , and  $\sigma_{k,i}^2$  is the corresponding variance. Learning of the Naive Bayes classifiers consists in estimating the parameters  $\{\mu_{k,i}, \sigma_{k,i}; i = 1, 2, \dots, p; k = 1, 2, \dots, K\}$ . One method for parameters estimation is called *maximum likelihood*.

#### 2.5.6. Logistic classifier

The *logistic classifier* is a supervised learning machine, which is based on a conditional probability function  $P(y=1|\mathbf{x}, \omega) = \frac{1}{1+e^{-\omega^T \mathbf{x}}} = h_\omega(\mathbf{x})$  and its complement  $P(y=0|\mathbf{x}, \omega) = 1 - h_\omega(\mathbf{x})$ , where  $\mathbf{x}$  is a feature vector and  $\omega$  is the set of parameters to be estimated. In general,  $P(y|\mathbf{x}, \omega) = h_\omega(\mathbf{x})^y (1 - h_\omega(\mathbf{x}))^{1-y}$ . To train a logistic classifier, a cost function, to be maximized, is defined according to  $L(\omega) = \prod_{i=1}^N h_\omega(\mathbf{x}_i)^{y_i} (1 - h_\omega(\mathbf{x}_i))^{1-y_i}$ , where  $\mathbf{x}_i$  is the  $i$ th feature vector,  $y_i$  specifies the class (0 or 1) for the  $i$ th observation, and  $N$  is the total number of feature vectors.

#### 2.5.7. Bayesian networks

A *Bayesian network* is a directed acyclic graph where each node represents a feature  $x_i$ . At each node, a conditional probability  $P(x_i|S_i)$  defines the relationship between the node (feature) and a subset of its parent nodes  $S_i$ , as it is shown in Fig. 8. The definition of a Bayesian network requires the computation of the marginal probabilities of the root nodes  $P(y)$  and the conditional probabilities of the non-root nodes given their parents, such as  $P(x_1|y)$  and  $P(x_3|y)$ . The network in Fig. 8 corresponds to an example where the binary variable at the root,  $y$ , represents the outcome of diagnosing Parkinson's disease ( $y = 1$  stands for TRUE while  $y = 0$  stands for FALSE). Given observed feature values (evidence) at some nodes, it is possible to compute the conditional probability of any other node such as  $P(y=1|x_1, x_2, x_3)$ , which is the probability of a positive Parkinson's disease diagnosis given the three feature values  $x_1, x_2$  and  $x_3$ . Each conditional probability function  $P(x_i|y)$  can be modeled as a Gaussian function.



**Fig. 8.** Example of a Bayesian network for Parkinson's disease detection.

**Table 1**

Settings of four sets of experiments to obtain regions of interest. The number of determined regions of interest is shown for each set.

MRI scanner	Male Population	Female Population
1.5 T	Set 1: 78 PD patients 23 controls 10 ROIs	Set 2: 36 PD patients 26 controls 12 ROIs
	Set 3: 153 PD patients 72 controls 12 ROIs	Set 4: 81 PD patients 38 controls 24 ROIs

## 3. Results

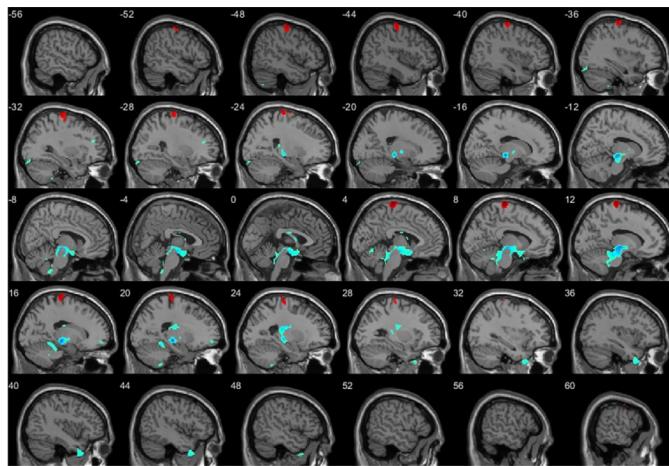
Four sets of experiments were conducted to obtain regions of interest with VBM. VBM was implemented by using two MATLAB toolboxes, the *Computational Anatomy Toolbox* (CAT 12) and the *Statistical Parametric Mapping* (SPM 12) toolbox. For each set of experiments, two templates were generated to represent two groups, a group of PD patients and another from controls. There were two sets of experiments where the MRI scans were generated with a 3-T scanner while the other two cases were based on MR images obtained with a 1.5-T scanner. A population of men was under study in two sets of experiments while a female population was the focus of the other two cases. Table 1 shows the settings and important results for the four sets of experiments to obtain regions of interest for further PD detection.

### 3.1. Regions of interest

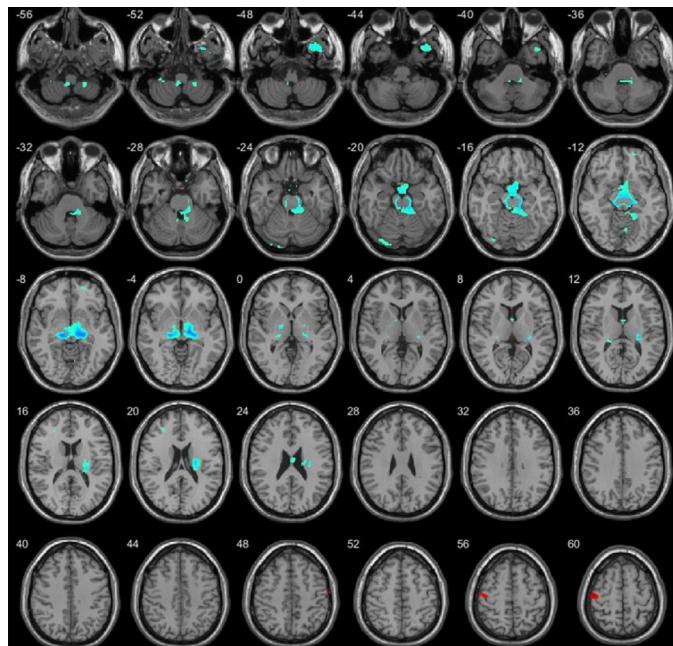
By using VBM, a template was generated for controls, and it was compared against the corresponding template for the group of PD patients. The result of this comparison is the *t-map*. Then, the *p*-value is used to determine those regions of interest, where the differences in gray matter concentration, between the two groups, are significant. The number of connected voxels is an additional requirement to obtain regions of interest. We considered 50 connected voxels to determine regions of interest. Table 1 specifies the number of regions of interest resulting from each experiment.

The results of VBM, for the case of a male population (231 patients and 95 controls) on 1.5-T and 3-T MR images, are shown in Figs. 9 and 10. These results are visualized for the cases of sagittal and axial views.

The two templates, from a group of 64 female controls and 117 PD female patients, were also compared by using VBM, and the results for sagittal and axial views are shown in Figs. 11 and 12, respectively.



**Fig. 9.** Color map of the sagittal view of the results of VBM applied to two-sample  $t$ -test from two groups. The first is conformed by 226 male PD patients and the second is obtained from 86 control male patients. The  $p$ -value was selected as 0.001. Red voxels correspond to  $t$ -values and blue voxels correspond to negative  $t$ -values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

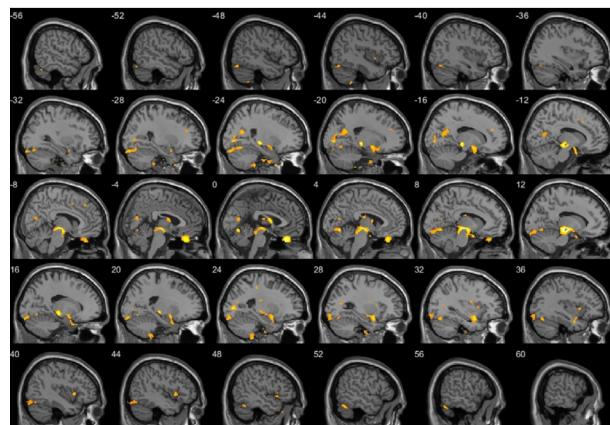


**Fig. 10.** Color map of the transverse section of the results of VBM and two-sample  $t$ -test from two groups. The first group is conformed by 226 male PD patients and the second from 86 control male patients. The  $p$ -value was selected as 0.001. Red voxels correspond to a positive  $t$ -value and blue voxels correspond to a negative  $t$ -value. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

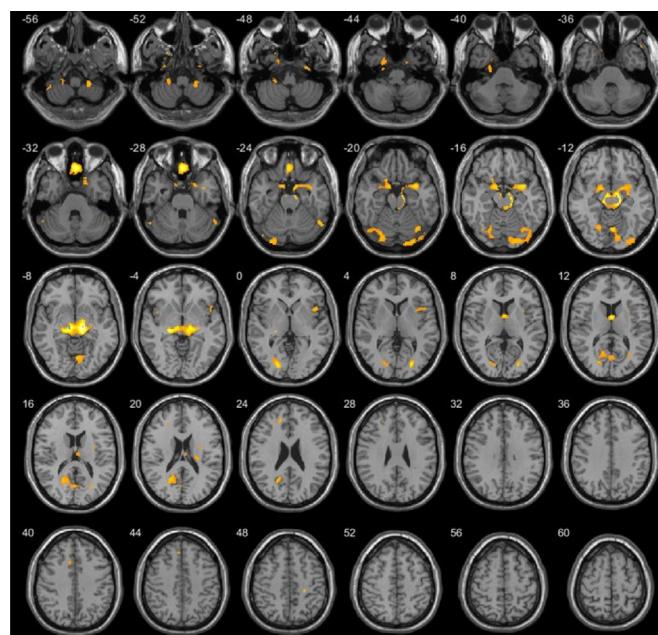
A group of 33 prodromal male PD patients was compared with the group of 95 male controls, and the results of VBM are shown in Figs. 13 (axial) and 14 (coronal). The motivation for carrying out this experiment was to find out about damaged regions at early stages.

### 3.2. PD detection results

Four sets of experiments were conducted to assess PD detection by using multiple classifiers. The setting for each experiment is shown in Table 1. At each experiment, seven classifiers were used, Logistic Classifier, Naive Bayes (NB), Bayesian Network (BN), Random



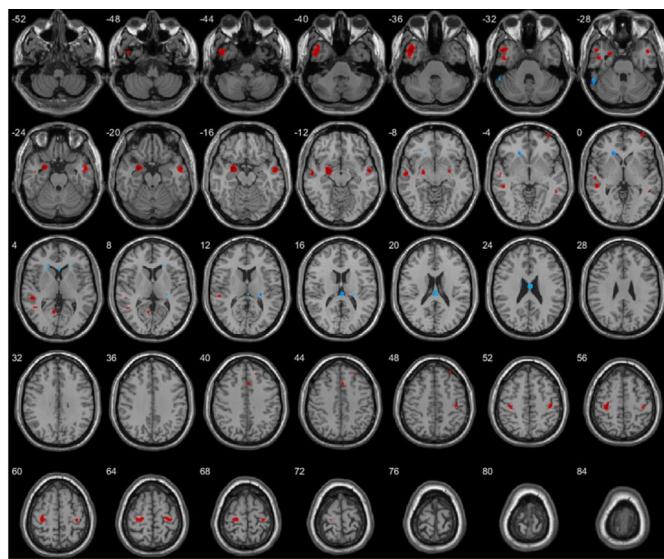
**Fig. 11.** Color map of the sagittal section of the results of VBM applied to two-sample  $t$ -tests from two groups. The first is conformed by 104 female PD patients and the second is obtained from 64 control female patients. The  $p$ -value was selected as 0.001. Only negative  $t$ -value voxels were found, and are rendered in yellow. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



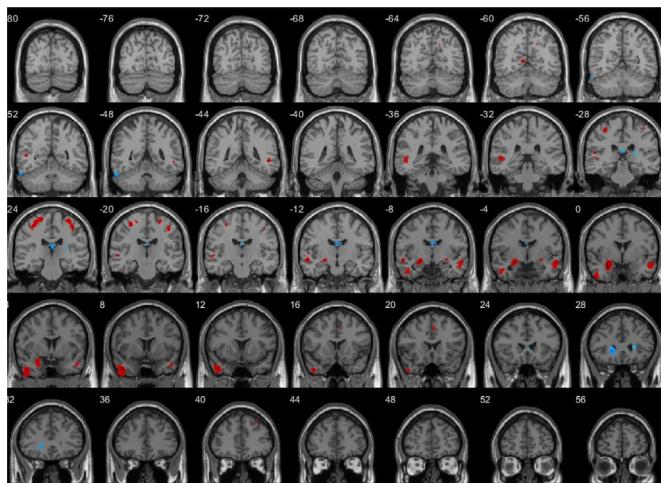
**Fig. 12.** Color map of the axial view of the results of VBM applied to two-sample  $t$ -tests from two groups. The first is conformed by 104 female PD patients and the second contains 64 control female patients. The  $p$ -value was selected as 0.001. Only negative  $t$ -value voxels were found, and are rendered in yellow. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Forest (RF), k-Nearest Neighbors (kNN), Multi-Layer Perceptron (MLP), Support Vector Machine (SVM). Feature extraction was implemented in JULIA 1.4. Feature selection, classification and performance assessment were implemented in WEKA.

Before classification and performance assessment, feature subset selection was accomplished by first applying PCA, and then running wrappers with the classifier to be tested, so that a very reduced set of features is obtained to evaluate the classifier detection performance. PCA and Wrappers considerably reduce the dimension of feature vectors. Table 2 shows the number of extracted features for each experiment described in Table 1. Besides the number of extracted features, Table 2 also presents the number of selected features after applying PCA, and the number of selected features after running wrappers with each classifier.



**Fig. 13.** Color map of the axial view of the results of VBM applied to two-sample *t*-tests from two groups. The first is conformed by 33 prodromal male PD patients and the second contains 95 male controls. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 14.** Color map of the coronal view of the results of VBM applied to two-sample *t*-tests from two groups. One group is conformed of 33 prodromal male PD patients and the other contains 95 male controls. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Table 2**

Number of extracted features from each set of experiments and number of selected features after applying PCA and after running wrappers with each classifier.

Number of features	Classifier	Male		Female	
		1.5 T	3 T	1.5 T	3 T
Extracted		2660	3192	3192	6384
Selected by PCA		1582	1911	1916	2180
Selected by Wrappers	Logistic	19	30	7	16
	RF	3	4	7	6
	NB	21	24	9	10
	BN	8	6	6	4
	kNN	9	8	9	2
	MLP	6	3	9	6
	SVM	6	6		

After feature subset selection, each classifier was tested by using *k-fold cross validation*. The training/testing set of feature vectors is partitioned into *k* sub-sets or *folds*,  $\{f_1, f_2 \dots, f_k\}$ . Then,  $k-1$  folds are used to train a classifier, while the remaining fold is used to measure classification performance. The process is repeated *k* times so that each fold is used for testing while different performance metrics are measured. Finally, the global performance metric is obtained by averaging the metrics at each fold. For this work,  $k = 10$ . During validation, if a feature vector, from a PD patient, is correctly classified, it is counted as a *True Positive* (TP); otherwise, it is a *False Negative* (FN). If a control instance is classified as such, it is counted as a *True Negative* (TN), otherwise it is a *False Positive* (FP). The performance metrics used are  $\text{accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$ ,  $\text{sensitivity} = \frac{TP}{TP+FN}$ ,  $\text{specificity} = \frac{TN}{TN+FP}$ , and  $\text{precision} = \frac{TP}{TP+FP}$ .

**Table 3** shows the performance results obtained by using seven different classifiers on two populations (men and woman) by using a 1.5-T MRI scanner. The results obtained with 3-T are shown in **Table 4**. The best results are highlighted in boldface in both Tables. There were two experiments where the SVM classifier did not achieve an accuracy above 70% (one in men and the other in women) and that is why these two cases are not shown in the Tables. The best detection performance achieved in men was 99.01% of accuracy (with a Naive Bayes classifier and 1.5-T scanner), 99.35% of sensitivity (SVM and 3 T), 100% of specificity (Naive Bayes and 1.5 T) and 100% of precision (Naive Bayes and 1.5 T). The best detection performance achieved in women is 96.97% of accuracy (Logistic classifier and 1.5 T), 100% of sensitivity (Bayesian Network and 1.5 T), 96.15% of specificity (Logistic classifier, MLP and 1.5 T), and 97.22% of precision (Logistic classifier and 1.5 T).

**Table 5** shows the selected features for the three best-performing classifiers in a male population for the case of 1.5-T and 3-T MRI scans. The corresponding selected features, after running wrappers with the three best classifiers in women, is shown in **Table 6**. The notation used to describe second-order statistics features consists of three parameters, the *direction* of the scans (*x*, *y*, *z*) that corresponds to three *MRI scan views* (coronal, sagittal, axial), the *angle* for generation of the co-occurrence matrix ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ), and the *region of interest*.

### 3.3. Regions with the most selected features

During classification of one subject as a PD patient or healthy individual, different features were extracted from different ROIs within this subject's MR images. It was found that some ROIs (obtained with VBM) were not used at all for classification of MR images during PD detection. Another finding is that different ROIs present different levels of contribution for PD detection. The higher the number of features, extracted from a particular ROI, the higher the contribution of this ROI to PD detection. **Table 5** shows the sets of selected features for the three best performing classifiers in a male population. **Table 5** also specifies the ROIs from which each feature was extracted. The corresponding sets of selected features, for the three best performing classifiers in women, are shown in **Table 6**. **Table 7** shows the ROIs used by the two best performing classifiers (in terms of accuracy) for each set of experiments in men (1.5 T and 3 T). **Table 7** specifies the number of times that extracted features were selected from each used ROI (second and fourth columns). The most used ROIs, for classification of features, are highlighted. **Table 8** shows the corresponding results for women. The most relevant ROIs, in terms of the number of features extracted, are highlighted in boldface.

The ROIs with the most selected features on 1.5-T MRI scans from a male population (**Table 7**) are shown in **Fig. 15**. ROI 4 is shown in red color and it is located in the Thalamus, while ROI 6 is shown in green and it is located in the right and left ventricles. According to the three views, both regions lie in the center of the

**Table 3**

PD detection results for male and female populations by using seven classifiers and 1.5-T MRI scans.

Classifier	Population	Accuracy	Sensitivity	Specificity	Precision
Logistic	Male	0.9604	0.9744	0.9130	0.9744
	Female	<b>0.9677</b>	0.9722	<b>0.9615</b>	<b>0.9722</b>
RF	Male	0.9208	0.9744	0.7391	0.9268
	Female	0.8710	0.9722	0.7308	0.8333
Naive Bayes	Male	<b>0.9901</b>	<b>0.9872</b>	<b>1</b>	<b>1</b>
	Female	0.8710	0.8889	0.8462	0.8889
Bayesian Net.	Male	0.9307	0.9487	0.8696	0.9610
	Female	0.8871	<b>1</b>	0.7308	0.8372
kNN	Male	0.9307	0.9744	0.7826	0.9383
	Female	0.9032	0.9444	0.8462	0.8947
MLP	Male	0.9802	<b>0.9872</b>	0.9565	0.9872
	Female	0.9355	0.9167	<b>0.9615</b>	0.9706
SVM	Female	0.8548	0.9167	0.7692	0.8462

**Table 4**

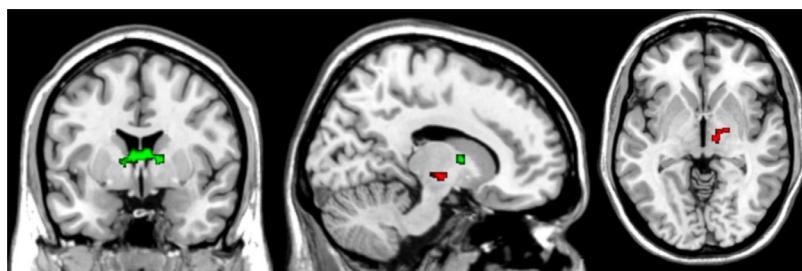
PD detection results for male and female populations by using seven classifiers and 3-T MRI scans.

Classifier	Population	Accuracy	Sensitivity	Specificity	Precision
Logistic	Male	0.9156	0.9281	<b>0.8889</b>	<b>0.9467</b>
	Female	<b>0.9328</b>	0.9506	<b>0.8947</b>	<b>0.9506</b>
RF	Male	0.8444	0.9216	0.6806	0.8598
	Female	0.7899	0.9136	0.5263	0.8043
Naive Bayes	Male	0.84	0.8431	0.8333	0.9149
	Female	0.8403	0.9012	0.7105	0.8690
Bayesian Net.	Male	0.7467	0.8366	0.5555	0.8
	Female	0.7059	0.8889	0.3158	0.7347
kNN	Male	0.8133	0.8627	0.7083	0.8627
	Female	0.7395	0.8519	0.5	0.7845
MLP	Male	0.8622	0.9346	0.7083	0.8720
	Female	0.8571	<b>0.9806</b>	0.6579	0.8556
SVM	Male	<b>0.9556</b>	<b>0.9935</b>	0.8750	0.9441

**Table 5**

Features selected by running wrappers with the three best performing classifiers for a male population.

Classifier	Selected features
Male Population and 1.5-T MRI scans	
Logistic	Entropy ( $y$ , 45°, ROI 4 and 5), ( $y$ , 0°, ROI 10, ( $y$ , 90°, ROI 7); difference entropy ( $x$ , 90°, ROIs 4 and 8), ( $x$ , 135°, ROI 6); dissimilarity ( $y$ , 90°, ROI 4), ( $x$ , 0°, ROIs 2 and 4); angular second moment ( $y$ , 0°, ROI 2), ( $x$ , 0°, ROI 6); cluster prominence ( $z$ , 135°, ROI 1 and 2); sum of squares ( $x$ , 135°, ROI 7); energy ( $x$ , 45°, ROI 5); histogram mean ROI 6; cluster shade ( $x$ , 0°, ROI 3); correlation 2 ( $z$ , 135°, ROI 5).
Naive Bayes	Angular second moment ( $y$ , 45°, ROI 8), ( $y$ , 90°, ROI 6), ( $y$ , 0°, ROI 3), ( $x$ , 135°, ROI 10); dissimilarity ( $y$ , 45°, ROI 7 and 8); sum average ( $y$ , 45°, ROI 6), ( $z$ , 0°, ROI 4), ( $z$ , 135°, ROI 6); sum of squares ( $x$ , 135°, ROI 7); correlation ( $x$ , 0°, ROI 4), correlation 1 ( $x$ , 135°, ROI 2); correlation 2 ( $z$ , 135°, ROI 9); histogram central moment 10 in ROI 9.
MLP	Entropy ( $y$ , 45°, ROI 4), ( $y$ , 0°, ROI 2); auto-correlation ( $x$ , 135°, ROI 6); inverse difference ( $x$ , 45°, ROI 10); sum of squares ( $x$ , 135°, ROI 7); sum average ( $z$ , 0°, ROI 4); histogram standard deviation in ROI 4.
Male Population and 3-T MRI scans	
Logistic	Entropy ( $x$ , 0°, ROI 12), ( $z$ , 135°, ROI 2 and 7); sum variance ( $z$ , 0°, ROI 1 and 8), ( $x$ , 90°, ROI 2); dissimilarity ( $x$ , 90°, ROI 6), ( $x$ , 135°, ROI 10); inverse difference ( $x$ , 90°, ROI 7 and 10), ( $y$ , 135°, ROI 1); correlation 1 ( $x$ , 135°, ROI 6), ( $z$ , 135°, ROI 6), ( $x$ , 0°, ROI 12); auto-correlation ( $z$ , 0°, ROI 2), ( $y$ , 90°, ROI 5); histogram central moment 3 in ROI 4 and 16 in ROI 7; maximum probability ( $y$ , 0°, ROI 2), ( $z$ , 90°, ROI 11); homogeneity ( $x$ , 0°, ROI 12); correlation ( $y$ , 135°, ROI 9); sum of squares ( $z$ , 0°, ROI 12).
Random Forest	Difference entropy ( $z$ , 45°, ROI 1); inverse difference ( $x$ , 90°, ROI 7); auto-correlation ( $x$ , 0°, ROI 4); histogram mean in ROI 12.
MLP	Entropy ( $z$ , 135°, ROI 2); inverse difference ( $x$ , 90°, ROI 7), ( $y$ , 0°, ROI 11), ( $z$ , 135°, ROI 2); sum average ( $x$ , 45°, ROIs 2 and 11), ( $y$ , 45°, ROI 2); auto-correlation ( $y$ , 135°, ROI 9); contrast ( $x$ , 0°, ROI 10), ( $y$ , 90°, ROI 6); correlation 1 ( $x$ , 45°, ROI 7), histogram central moment 2 in ROI 12.



**Fig. 15.** The two ROIs with the most selected features on 1.5-T MRI scans on a male population.

**Table 6**

Features selected by running wrappers with the three best performing classifiers for a female population.

Classifier	Selected features
Female Population and 1.5-T MRI scans	
Logistic	Entropy ( $y$ , $0^\circ$ , ROI 7); difference variance ( $x$ , $135^\circ$ , ROI 11), ( $x$ , $135^\circ$ , ROI 6); difference entropy ( $z$ , $45^\circ$ , ROI 1); angular second moment ( $y$ , $135^\circ$ , ROI 6); cluster prominence ( $x$ , $90^\circ$ , ROI 3); autocorrelation ( $y$ , $90^\circ$ , ROI 10).
kNN	Energy ( $z$ , $0^\circ$ , ROI 11); auto-correlation ( $x$ , $45^\circ$ , ROI 10), ( $y$ , $90^\circ$ , ROI 10); angular second moment ( $z$ , $135^\circ$ , ROI 4), ( $z$ , $90^\circ$ , ROI 8); inverse difference ( $y$ , $45^\circ$ , ROI 2); sum average ( $z$ , $45^\circ$ , ROI 8); cluster prominence ( $z$ , $45^\circ$ , ROI 4); correlation 1 ( $y$ , $90^\circ$ , ROI 1).
MLP	Entropy ( $z$ , $135^\circ$ , ROI 5); sum average ( $y$ , $90^\circ$ , ROIs 2 and 3), ( $z$ , $45^\circ$ , ROI 8); sum entropy ( $x$ , $45^\circ$ , ROI 10); auto-correlation ( $x$ , $90^\circ$ , ROI 6); dissimilarity ( $x$ , $135^\circ$ , ROI 7); histogram central moment 18 in ROI 7.
Female Population and 3-T MRI scans	
Logistic	Dissimilarity ( $x$ , $0^\circ$ , ROI 13), ( $y$ , $0^\circ$ , ROI 12), ( $y$ , $45^\circ$ , ROI 14); contrast ( $z$ , $90^\circ$ , ROI 12 and 19), ( $x$ , $45^\circ$ , ROI 21), ( $x$ , $90^\circ$ , ROI 1); maximum probability ( $z$ , $45^\circ$ , ROI 2 and 8); sum variance ( $x$ , $90^\circ$ , ROI 18), ( $z$ , $90^\circ$ , ROI 3); correlation ( $z$ , $0^\circ$ , ROI 16), ( $z$ , $45^\circ$ , ROI 19), ( $z$ , $135^\circ$ , ROI 19); difference entropy ( $x$ , $0^\circ$ , ROI 21); sum average ( $y$ , $0^\circ$ , ROI 1).
Naive Bayes	Correlation 1 ( $y$ , $90^\circ$ , ROI 7), ( $y$ , $135^\circ$ , ROI 16); correlation 2 ( $x$ , $0^\circ$ , ROIs 20 and 21); correlation ( $z$ , $90^\circ$ , ROI 4); auto-correlation ( $x$ , $45^\circ$ , ROI 22); sum average ( $y$ , $0^\circ$ , ROI 1), ( $z$ , $45^\circ$ , ROI 6); entropy ( $y$ , $135^\circ$ , ROI 13); energy ( $y$ , $90^\circ$ , ROI 24); contrast ( $y$ , $90^\circ$ , ROI 17); difference variance ( $y$ , $45^\circ$ , ROI 13); homogeneity ( $y$ , $135^\circ$ , ROI 10); sum variance ( $x$ , $45^\circ$ , ROI 12).
MLP	Correlation 1 ( $y$ , $45^\circ$ , ROI 19), ( $y$ , $45^\circ$ , ROI 24); correlation 2 ( $x$ , $135^\circ$ , ROI 7); cluster prominence ( $x$ , $135^\circ$ , ROI 7); contrast ( $z$ , $90^\circ$ , ROI 20); homogeneity ( $y$ , $135^\circ$ , ROI 12).

**Table 7**

The regions of interest with the most selected features by the two best performing classifiers for each set of experiments in a male population. It is specified the number of times that a region of interest is used to select features for classification.

Men, 1.5-T scans, Naive Bayes and MLP		Men, 3-T scans, Logistic and MLP	
ROI	Number of Selected Features	Number of ROIs	Number of Selected Features
2	1	1	2
3	1	2	<b>8</b> (22.22%)
4	<b>4</b> (18.18%)	4	1
6	<b>6</b> (27.27%)	5	1
7	3	6	4
8	2	7	<b>5</b> (13.89%)
9	2	8	1
10	3	9	2
		10	3
		11	3
		12	<b>6</b> (16.67%)
Total:	22	Total:	36

**Table 8**

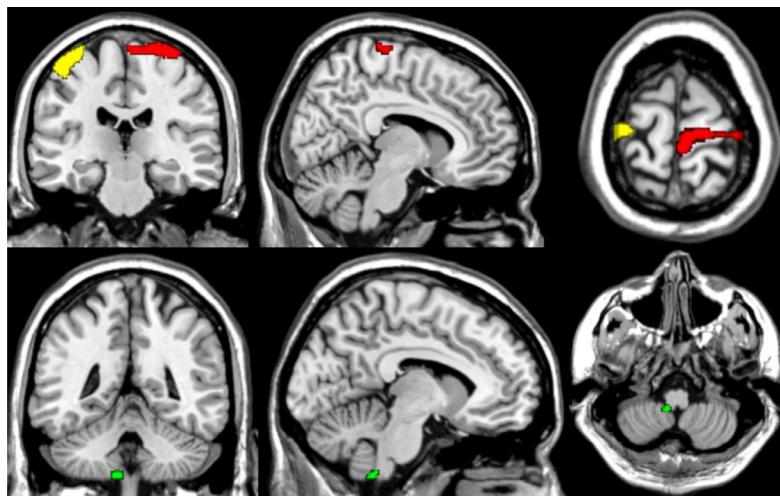
The regions of interest with the most selected features by the two best performing classifiers for each set of experiments in women. It is specified the number of times that a region of interest is used to select features for classification.

Women, 1.5-T scans, Logistic and MLP		Women, 3-T scans, Logistic and MLP	
ROI	Number of Selected Features	Number of ROIs	Number of Selected Features
1	1	1	<b>3</b> (9.38%)
2	1	2	2
3	2	3	1
5	2	6	1
6	<b>3</b> (15.79%)	7	1
7	<b>3</b> (15.79%)	8	2
8	2	10	1
10	2	12	2
11	1	13	1
18	2	14	<b>3</b> (9.38%)
		15	1
		16	1
		17	1
		18	<b>4</b> (12.5%)
		19	<b>4</b> (12.5%)
		20	2
		21	2
Total:	19	Total:	32

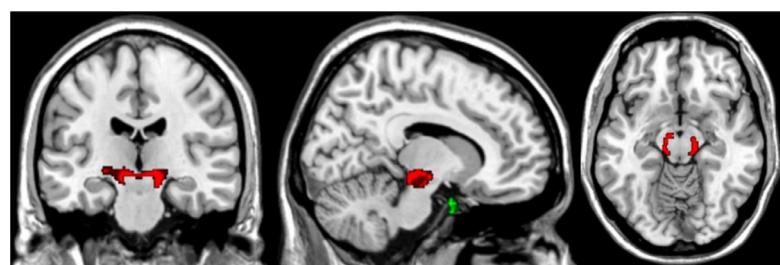
brain. The three ROIs with the most selected features, for the case of men and 3-T MRI scans, are shown in Fig. 16. In this case, the need for two rows is due to the fact that the regions lie at different depths so that it is not possible to visualize all the ROIs on the same planes. ROI 2 is red (first row) and it comprises the union of the two hemispheres at the cortex. ROI 7 is green (second row)

and it lies at the inferior part of the cerebellum. ROI 14 is yellow (first row) and it is located in the cortex

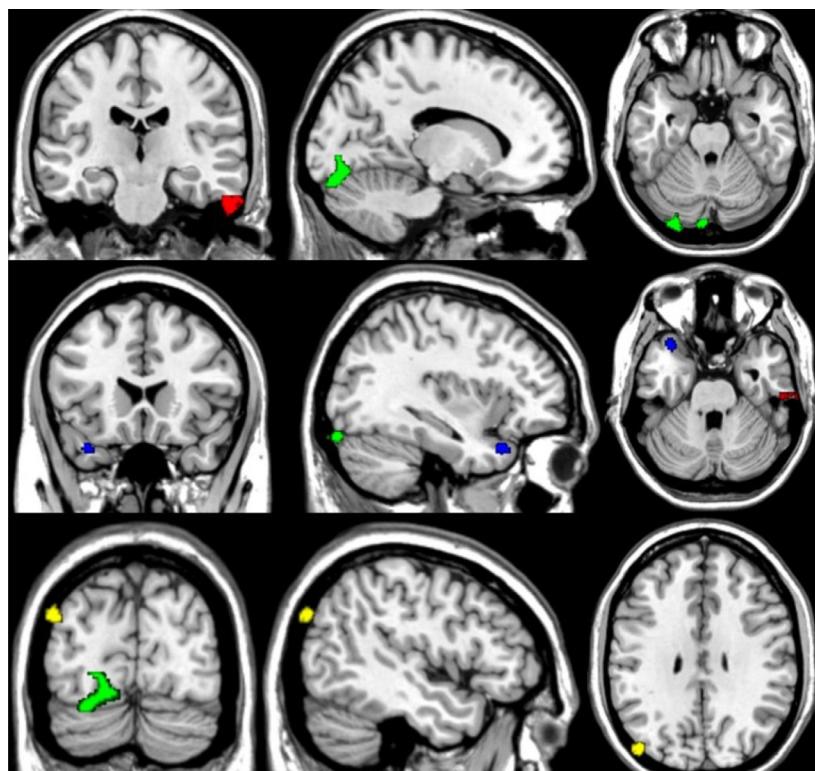
By observing Table 8, the two ROIs with the most selected features, for women and 1.5-T MRI scans, are illustrated in Fig. 17. ROI 6 is shown in red and it corresponds to the thalamus, while ROI 7 is green and corresponds to the brainstem. Fig. 18 shows the four most relevant ROIs in women on 3-T MRI scans. ROI 1 is shown in



**Fig. 16.** The three ROIs with the most selected features on 3-T MRI scans from a male population. ROIs 2 (red) and 14 (yellow) are shown in the first row while ROI 7 (green) is located in the second row. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 17.** The two ROIs with the most selected features for women on 1.5-T MRI scans.



**Fig. 18.** The four ROIs with the most selected features for women on 3-T MRI scans. ROIs 1 (red) and 14 (green) are shown in the first row. ROIs 18 (blue) and 19 (yellow) are located in the second and third rows, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

red color (first row) and it is found in the inferior temporal gyrus. ROI 14 is green (first row), and it is located in the visual cortex, above the cerebellum. ROI 18 is blue (second row) and it lies in the temporal gyrus. ROI 19 is yellow (third row) and it is in the cortex.

## 4. Discussion

### 4.1. Dataset

In this study, we took advantage of the number of PD patients and controls on the PPMI MRI dataset. In previous works, experiments were not conducted in terms of gender due to the low number of subjects in other datasets. According to [Table 1](#), one weakness of the PPMI MRI dataset is that the number of female subjects is almost half the number of men, which reflects the situation of men being more prone to PD.

### 4.2. Regions of interest in men and women

[Figs. 9](#) and [10](#) show that the areas more affected in men are the basal ganglia, the brainstem, fourth ventricle, lateral ventricle and cerebellum. The cerebellum, frontal lobe, temporal lobe and cortex are also included. [Figs. 11](#) and [12](#) show that the areas affected in women are the occipital lobe, thalamus basal ganglia, a small part of the cerebellum, and the frontal lobe.

By comparing [Fig. 9](#) with [Fig. 11](#) as well as [Fig. 10](#) with [Fig. 12](#), it is observed that regions of interest in men are bigger than those in women, a finding that somehow agrees with medical facts which state that more men than women are diagnosed with PD by almost a 2 to 1 margin. In the other hand, according to [Table 1](#), the number of resulting regions of interest, after VBM, are more in women (12 ROIs in 1.5-T MRI scans and 24 ROIs in 3-T MRI scans) than in men (10 ROIs in 1.5-T MRI scans and 12 ROIs in 3-T MRI scans). An explanation for the last finding is that regions of interest, in men and women, are generally scattered over the same brain zones. This is to say that an area that includes multiple regions of interest, in men, is equivalent to that including more and smaller regions of interest for women. Furthermore, the number of selected features for classification in women is more reduced than in men, as it is shown in [Table 5](#) vs. [Table 6](#). The consequence of selecting a smaller number of features in women is a reduction in the number of used ROIs for them, where multiple ROIs are not used for classification.

By comparing a group of prodromal male PD patients with a group of prodromal male controls, as it is shown in [Figs. 13](#) and [14](#), it was found that the cortex is an area of interest at early stages.

### 4.3. Classification of PPMI MR images

The best classification results were obtained by mostly using a Naive Bayes classifier in men and a Logistic classifier in women. Almost all the highest metric values were obtained by using 1.5-T MRI scans.

A comparison of the best detection performance values, between men and women, does not show significant differences; however, we believe that the factors, associated to PD detection, are dependent on gender. According to all the conducted experiments in [Tables 3](#) and [4](#), the highest detection performance values were obtained for men, the group with the largest dataset since detection performance depends on the number of observations.

### 4.4. Features

The selected feature subsets arise from two different groups of features, *first-order statistics* and *second-order statistics*. According

to [Tables 5](#) and [6](#), it is observed that almost all the selected features arise from the second-order statistics on ROIs. This is obvious since the co-occurrence matrix conveys information regarding the way similar and different voxel intensity values are distributed spatially. The most relevant second-order statistics features are entropy and correlation.

### 4.5. Regions with the most selected features

[Fig. 15](#) shows a male region located around putamen and thalamus; brain regions where the generation of signals for involuntary movement and instincts (very related to characteristic PD symptoms) take place. As a contrast, according to the upper row in [Fig. 16](#), other regions, with the most selected features, lies on the upper cortex, a region where reasoning, decision making, and awareness take place. An explanation for the last statement is found in neurobiology, which claims that the female central nervous system is stronger than the male one so that men are more vulnerable to PD deterioration consequences.

For the case of women, [Fig. 17](#) shows regions around the thalamus and brainstem. By observing [Fig. 18](#), another affected region is the temporal gyrus, which is associated with visual stimuli processing, and spatial awareness. The detected cortex area is not as large as the case in men.

Most of the related work for automated PD detection is focused on searching for damages around the striatum region of the brain. Instead of addressing the problem in the same way and looking just for regions with dopamine loss, we looked for regions where there is a reduction of the tissue thickness, a difference between PD patients and control subjects. Besides the striated region, an automated and quantitative analysis should be conducted over the whole brain. It turned out that regions of interest were found in the cortex, a finding that agrees with the fact that the cortex is a region of interest at early stages since it was detected in prodromal patients.

### 4.6. Comparison with previous work

[Table 9](#) shows the performance of different methods, including the proposed one. The purpose of this Table is to show different methodologies and the corresponding results. This Table also shows the framework used by each method and the datasets used by the scientific community. It is observed that half of the methods are characterized by an accuracy above 90% and that the PPMI MRI dataset is the most extensive dataset in terms of the number of PD patients and healthy controls. It is observed that the proposed method is the most extensive in terms of the diversity of tested classifiers, in terms of conducting PD detection for separate populations of men and women, and in terms of testing multiple feature subsets to select the best classifier architecture. The proposed method and the one, from Esmaeilzadeh et al. [21], are the two works that achieve the highest accuracy. Two limitations of our work are that we have not reached 100% of accuracy and have not tried deep learning as a tool for PD detection so far.

### 4.7. Future direction

In this work, features arise from multiple arrays, one-dimensional arrays (vectors) corresponding to first-order histograms, and two-dimensional arrays corresponding to co-occurrence matrices. One vector and multiple matrices are obtained from each region. One co-occurrence matrix ( $256 \times 256$  array) is obtained for each plane (axial, sagittal, coronal) at one angle ( $0^\circ, 45^\circ, 90^\circ, 135^\circ$ ) at different depths (on  $x$  for coronal planes,  $y$  for sagittal planes, and  $z$  for axial planes). The number of co-occurrence matrices, for each ROI, is not small and that is the main

**Table 9**

A comparison with different PD detection methods.

Author and year	Dataset	Framework	Performance
Long et al. [18]	MRI from 19 PD patients and 27 healthy subjects	SVM and features extracted from rsfMRI	accuracy 86.96%, sensitivity 78.95%, specificity 92.59%
Lei et al. [19]	PPMI MRI dataset	Sparse feature selection framework	accuracy 86.48%
Sivaranjini et al. [20]	PPMI MRI dataset	CNN	accuracy 88.9%
Esmailzadeh et al. [21]	PPMI MRI dataset and personal information (age, gender)	Deep learning	accuracy 100%
Shah et al. [22]	PPMI MRI dataset	CNN	accuracy 93%
Slavatore et al. [23]	MRI from 28 PD patients and 28 healthy controls	SVM, PCA for feature extraction	accuracy, sensitivity and specificity above 90%
Shinde et al. [24]	Neuromelanine MRI	CNN	85% of accuracy
Amoroso et al. [25]	PPMI MRI dataset	SVM and clinical features	accuracy 93%, sensitivity 93%, specificity 92%
Proposed method	PPMI MRI dataset	Seven classifiers, Haralick features and wrapper feature selection	accuracy 99.01% (men) and 96.97% (women), sensitivity 99.35% (men) and 100% (women), specificity 100% (men) and 96.15% (women)

reason why the number of extracted features is huge and the need for further feature selection techniques.

To avoid increasing the number of extracted features, we limited feature extraction to first- and second-order statistics, and did not explore other techniques for generation of features such as frequency-based techniques (wavelet transform, Fourier transform).

Another strategy, to avoid the use of feature subset selection and the generation of a huge number of features, consists of directly feeding co-occurrence matrices and/or ROIs (fitted as tensors of gray level values) to *Convolutional Neural Networks* (CNN). A CNN is suitable for this application since multiple first- and second-order statistics histograms are high dimensional data and they can be modeled as three-dimensional tensors for processing (convolution) and classification. The use of CNNs, as an extension for classification of brain signals, has been possible by formatting fNIRS (*functional Near Infra Red Spectroscopy*) signals as second- and third-order tensors [38]. Furthermore, the main motivation will be the discovery of linear features, determined by the weights of the kernels at convolutional layers, and which might be more appropriate for classification of MR images to assist the diagnosis of PD.

## 5. Conclusion

In conclusion, the proposed method for classification of PPMI MR images with VBM, from controls and PD patients, provides high performance as a tool to assist in the diagnosis of PD. This study was conducted separately for male and female populations with good results in both cases.

The results show that different regions are useful to classify PPMI MR images depending on gender. The areas more affected in men are the basal ganglia, the brainstem, fourth ventricle, lateral ventricle and cerebellum. The corresponding regions for women are the occipital lobe, thalamus basal ganglia, a small part of the cerebellum, and the frontal lobe.

Regions of interest were detected by using VBM, whereas features were generated by obtaining multiple co-occurrence matrices from each region. Because of the huge amount of extracted features, PCA and subsequent feature subset selection considerably reduce the number of features.

The search for regions of interest is carried out over the brain as opposed to some related works, which were focused on the striatum region of the brain. Thus, other regions of interest are found such as the cortex, visual cortex and the temporal gyrus.

## Declaration of Competing Interest

None.

## Acknowledgments

The authors would like to acknowledge the support of the *National Council for Research and Technology* (CONACYT) in Mexico (Scholarship 934454 and stimulus 68150). We made use of a public dataset, which was collected by the *Michael J. Fox Foundation*.

## References

- [1] J. Michael, Fox Foundation for Parkinson Research, Parkinson's disease causes, 2018. (Retrieved from: <https://www.michaeljfox.org/understanding-parkinsons/living-with-pd.html>)
- [2] B. Heim, F. Krismer, R. De Marzi, K. Sppi, Magnetic resonance imaging for the diagnosis of Parkinson's disease, *J. Neural Transm.* 124 (8) (2017) 915–964, doi:[10.1007/s00702-017-1717-8](https://doi.org/10.1007/s00702-017-1717-8).
- [3] V. Fioravanti, F. Benuzzi, L. Codeluppi, S. Contardi, F. Cavallieri, P. Nichelli, F. Valzania, MRI correlates of Parkinson's disease progression: a voxel-based morphometry study, *Parkinson's Dis.* (2015) 1–8, doi:[10.1155/2015/378032](https://doi.org/10.1155/2015/378032).
- [4] A. Hotter, R. Esterhammer, M.F. Schocke, K. Seppi, Potential of advanced MR imaging techniques in the differential diagnosis of Parkinsonism, *Mov. Disord.* 24 (2009) 711–720, doi:[10.1002/mds.22648](https://doi.org/10.1002/mds.22648).
- [5] N.W. Sterling, G. Du, M.N. Lewis, Striatal shape in Parkinson's disease, *Neurobiol. Aging* 34 (11) (2011) 2510–2516, doi:[10.1016/j.neurobiolaging.2013.05.017](https://doi.org/10.1016/j.neurobiolaging.2013.05.017).
- [6] N. Pyatigorskaya, B. Magnin, M. Mongin, L. Yahia-Cherif, R. Valabregue, D. Arnaldi, C. Ewenzyk, C. Poupon, M. Vidailhet, S. Lehéricy, Comparative study of MRI biomarkers in the substantia nigra to discriminate idiopathic Parkinson disease, *Am. J. Neuroradiol.* 39 (8) (2018) 1460–1467, doi:[10.3174/ajnr.A5702](https://doi.org/10.3174/ajnr.A5702).
- [7] J. Ashburner, K.J. Friston, Voxel-based morphometry - the methods, *Neuroimage* 11 (6) (2000) 805–821, doi:[10.1006/nimg.2000.0582](https://doi.org/10.1006/nimg.2000.0582).
- [8] J. Ashburner, K.J. Friston, Voxel based morphometry, *Encycl. Neurosci.* 2009 (2009) 471–477, doi:[10.1016/B978-008045046-9.00306-5](https://doi.org/10.1016/B978-008045046-9.00306-5).
- [9] J.C. Blair, M.J. Barrett, J. Patric, J.L. Flanigan, S.A. Sperling, W.J. Elias, T.J. Druzgal, Brain MRI reveals ascending atrophy in Parkinson's disease across severity, *Front. Neurol.* (2019) 1–15, doi:[10.3389/fneur.2019.01329](https://doi.org/10.3389/fneur.2019.01329).
- [10] C. Summerfield, C. Junqué, E. Tolosa, P. Salgado-Pineda, B. Gómez-Ansón, M.J. Martí, P. Pastor, B. Ramírez-Ruiz, J. Mercader, Structural brain changes in Parkinson disease with dementia: a voxel-based morphometry study, *Arch. Neurol.* 62 (2) (2005) 281–285, doi:[10.1001/archneur.62.2.281](https://doi.org/10.1001/archneur.62.2.281).
- [11] L. Krajeovicova, P. Klobusikova, I. Rektorova, Gray matter changes in Parkinson's and Alzheimer's disease and relation to cognition, *Neuroimaging* 19 (85) (2019) 1–9, doi:[10.1007/s11910-019-1006-z](https://doi.org/10.1007/s11910-019-1006-z).
- [12] L.H. Mozley, R.C. Gur, D.P. Mozley, R.E. Gur, Striatal dopamine transporters and cognitive functioning in healthy men and women, *Am. J. Psychiatry* 158 (9) (2001) 1492–1499, doi:[10.1176/appi.ajp.158.9.1492](https://doi.org/10.1176/appi.ajp.158.9.1492).
- [13] I.N. Miller, A. Cronin-Golomb, Gender differences in Parkinson's disease: clinical characteristics and cognition, *Mov. Disord.* 25 (16) (2010) 2695–2703, doi:[10.1002/mds.23388](https://doi.org/10.1002/mds.23388).
- [14] C.A. Haaxma, B.R. Bloem, G.F. Borm, W.J.G. Oyen, K.L. Leenders, S. Eshuis, J. Booij, D.E. Dluzen, M.W.I.M. Horstink, Gender differences in Parkinson's disease, *J. Neurol. Neurosurg. Psychiatry* 78 (8) (2007) 819–824, doi:[10.1136/jnnp.2006.103788](https://doi.org/10.1136/jnnp.2006.103788).

- [15] P.H. Kuo, R. Avery, E. Krupinski, L. Hong, A. Bauer, S. Sherman, N. McMillan, J. Seibyl, G. Zubal, Receiver-operating-characteristic analysis of an automated program for analyzing striatal uptake of I-loflupane SPECT images: calibration using visual reads, *J. Nucl. Med. Technol.* 41 (1) (2013) 26–31, doi:[10.2967/jnmt.112.114827](https://doi.org/10.2967/jnmt.112.114827).
- [16] P.H. Kuo, L.H. Hong, R. Avery, E. Krupinski, A. Bauer, S. Sherman, N. McMillan, J. Seibyl, G. Zubal, Evaluation of an objective striatal analysis program for determining laterality in uptake of I-loflupane SPECT: comparison to clinical symptoms and to visual reads, *J. Nucl. Med. Technol.* 42 (1) (2014) 1–4, doi:[10.2967/jnmt.113.134940](https://doi.org/10.2967/jnmt.113.134940).
- [17] H.D. Tagare, C. DeLorenzo, S. Chelikani, R.K. Fulbright, Voxel-based logistic analysis of PPMI control and Parkinson's disease DaTscans, *Neuroimage* 152 (2017) 299–311, doi:[10.1016/j.neuroimage.2017.02.067](https://doi.org/10.1016/j.neuroimage.2017.02.067).
- [18] D. Long, J. Wang, M. Xuan, Q. Gu, X. Xu, D. Kong, M. Zhang, Automatic classification of early Parkinson's disease with multi-modal MR Imaging, *Plos One* (2012) 1–9, doi:[10.1371/journal.pone.0047714](https://doi.org/10.1371/journal.pone.0047714).
- [19] H. Lei, Y. Zhao, Y. Wen, Q. Luo, Y. Cai, G. Liu, B. Lei, Sparse feature learning for multi-class Parkinson's disease classification, *Technol. Health Care* 26 (1) (2018) 193–203, doi:[10.3233/thc-174548](https://doi.org/10.3233/thc-174548).
- [20] S. Sivaranjini, C.M. Sujatha, Deep learning based diagnosis of Parkinson's disease using convolutional neural network, *Multimed. Tools Appl.* 79 (2020) 15467–15479, doi:[10.1007/s11042-019-7469-8](https://doi.org/10.1007/s11042-019-7469-8).
- [21] S. Esmaeilzadeh, Y. Yang, E. Adeli, End-to-end Parkinson disease diagnosis using brain MR-images by 3D-CNN, *Comput. Vis. Pattern Recognit.* (2018) 1–7.
- [22] P.M. Shah, A. Zeb, U. Shah, S.F.A. Zaidi, M.A. Shah, Detection of Parkinson disease in brain MRI using convolutional neural network, in: 2018 24th International Conference on Automation and Computing (ICAC), 2018, pp. 1–6.
- [23] C. Salvatore, A. Cerasa, I. Castiglioni, F. Gallivanone, A. Augimeri, M. Lopez, G. Arabia, M. Morelli, M.C. Gilardi, A. Quattrone, Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and progressive supranuclear palsy, *J. Neurosci. Methods* 222 (2014) 230–237, doi:[10.1016/j.jneumeth.2013.11.016](https://doi.org/10.1016/j.jneumeth.2013.11.016).
- [24] S. Shinde, S. Prasad, Y. Saboo, R. Kaushick, J. Saini, P.K. Pal, M. Ingalhalikar, Predictive markers for Parkinson's disease using deep neural nets on neuromelanin sensitive MRI, *Neuroimage* 22 (2019) 101748, doi:[10.1016/j.nicl.2019.101748](https://doi.org/10.1016/j.nicl.2019.101748).
- [25] N. Amoroso, M. La Rocca, A. Monaco, R. Bellotti, S. Tangaro, Complex networks reveal early MRI markers of Parkinson's disease, *Med. Image Anal.* 48 (2018) 12–24, doi:[10.1016/j.media.2018.05.004](https://doi.org/10.1016/j.media.2018.05.004).
- [26] G. Solana-Lavalle, J.C. Galán-Hernández, R. Rosas-Romero, Automatic Parkinson disease detection at early stages as pre-diagnosis tool by using classifiers and a small set of vocal features, *Biocybern. Biomed. Eng.* 40 (1) (2020) 505–516, doi:[10.1016/j.bbe.2020.01.003](https://doi.org/10.1016/j.bbe.2020.01.003).
- [27] C.O. Sakar, G. Serbes, A. Gunduz, H.C. Tunc, H. Nizam, B.E. Sakar, M. Tutuncu, T. Aydin, M.E. Isenkul, H. Apaydin, A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform, *Appl. Soft Comput.* 74 (2019) 255–263, doi:[10.1016/j.asoc.2018.10.022](https://doi.org/10.1016/j.asoc.2018.10.022).
- [28] S.A. Mostafa, A. Mustapha, M.A. Mohammed, R.I. Hamed, N. Arunkumar, M.K. Abd Ghani, M.M. Jaber, S.H. Khaleefah, Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease, *Cognit. Syst. Res.* 54 (2019) 90–99, doi:[10.1016/j.cogsys.2018.12.004](https://doi.org/10.1016/j.cogsys.2018.12.004).
- [29] D. Gupta, A. Julka, S. Jain, T. Aggarwal, A. Khanna, N. Arunkumar, V.H.C. de Albuquerque, Optimized cuttlefish algorithm for diagnosis of Parkinson's disease, *Cognit. Syst. Res.* 52 (2018) 36–48, doi:[10.1016/j.cogsys.2018.06.006](https://doi.org/10.1016/j.cogsys.2018.06.006).
- [30] S.A. Mostafa, A. Mustapha, S.H. Khaleefah, M.S. Ahmad, M.A. Mohammed, Evaluating the performance of three classification methods in diagnosis of Parkinson's disease, in: International Conference on Soft Computing and Data Mining, 2018, pp. 43–52.
- [31] D. Gupta, S. Sundaram, A. Khanna, A.E. Hassanien, V.H.C. De Albuquerque, Improved diagnosis of Parkinson's disease using optimized crow search algorithm, *Comput. Electr. Eng.* 68 (2018) 412–424, doi:[10.1016/j.compeleceng.2018.04.014](https://doi.org/10.1016/j.compeleceng.2018.04.014).
- [32] C.R. Pereira, D.R. Pereira, S.A.T. Weber, C. Hook, V.H.C. de Albuquerque, J.P. Papa, A survey on computer-assisted Parkinson's disease diagnosis, *Artif. Intell. Med.* 95 (2019) 48–63, doi:[10.1016/j.artmed.2018.08.007](https://doi.org/10.1016/j.artmed.2018.08.007).
- [33] S.K. Yadav, N. Kathiresan, S. Mohan, G. Vasileiou, A. Singh, D. Kaura, E.R. Melhem, R.K. Gupta, E. Wang, F.M. Marincola, Gender-based analysis of cortical thickness and structural connectivity in Parkinson's disease, *J. Neurol.* 263 (11) (2016) 2308–2318, doi:[10.1007/s00415-016-8265-2](https://doi.org/10.1007/s00415-016-8265-2).
- [34] R.M. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, *IEEE Trans. Syst. Man Cybern.* 3 (6) (1973) 610–621.
- [35] T. Lofstedt, P. Brynolfsson, T. Asklund, T. Nyholm, A. Garpebring, Gray-level invariant Haralick texture features, *Plos One* 14 (2) (2019) 1–18, doi:[10.1371/journal.pone.0212110](https://doi.org/10.1371/journal.pone.0212110).
- [36] R. Kohavi, G.H. John, Wrappers for feature subset selection, *Artif. Intell.* 97 (1997) 273–324.
- [37] F. Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain, *Psychol. Rev.* 65 (6) (1958) 386.
- [38] R. Rosas, E. Guevara, K. Peng, D.K. Nguyen, P. Lesage, E.W. Pouliot, E.W. Lima-Saad, Prediction of epileptic seizures with convolutional neural networks and functional near-infrared spectroscopy signals, *Comput. Biol. Med.* 111 (103355) (2019) 1–10, doi:[10.1016/j.combiomed.2019.103355](https://doi.org/10.1016/j.combiomed.2019.103355).