



Distinguishing Parkinson's disease from atypical parkinsonian syndromes using PET data and a computer system based on support vector machines and Bayesian networks

Fermín Segovia^{1*}, Ignacio A. Illán¹, Juan M. Górriz¹, Javier Ramírez¹, Axel Rominger² and Johannes Levin³

¹ Department of Signal Theory, Networking and Communications, University of Granada, Granada, Spain, ² Department of Nuclear Medicine, Ludwig Maximilian University of Munich, Munich, Germany, ³ Department of Neurology, University of Munich, Munich, Germany

OPEN ACCESS

Edited by:

John Suckling,
University of Cambridge, UK

Reviewed by:

Elmar W. Lang,
University of Regensburg, Germany
Jigar Jogia,
Aston University, UK

*Correspondence:

Fermín Segovia
fsegovia@ugr.es

Received: 31 July 2015

Accepted: 23 October 2015

Published: 05 November 2015

Citation:

Segovia F, Illán IA, Górriz JM, Ramírez J, Rominger A and Levin J (2015) Distinguishing Parkinson's disease from atypical parkinsonian syndromes using PET data and a computer system based on support vector machines and Bayesian networks.
Front. Comput. Neurosci. 9:137.
doi: 10.3389/fncom.2015.00137

Differentiating between Parkinson's disease (PD) and atypical parkinsonian syndromes (APS) is still a challenge, specially at early stages when the patients show similar symptoms. During last years, several computer systems have been proposed in order to improve the diagnosis of PD, but their accuracy is still limited. In this work we demonstrate a full automatic computer system to assist the diagnosis of PD using ¹⁸F-DMFP PET data. First, a few regions of interest are selected by means of a two-sample *t*-test. The accuracy of the selected regions to separate PD from APS patients is then computed using a support vector machine classifier. The accuracy values are finally used to train a Bayesian network that can be used to predict the class of new unseen data. This methodology was evaluated using a database with 87 neuroimages, achieving accuracy rates over 78%. A fair comparison with other similar approaches is also provided.

Keywords: Bayesian network, support vector machine, 18F-DMFP PET, Parkinson's disease, multivariate analysis

1. INTRODUCTION

One of the neuropathological hallmarks of Parkinson's disease (PD; Greenberg et al., 2012) is a substantial decrease in the dopamine content of the striatum due to the progressive death of dopaminergic neurons of the nigrostriatal pathway (a neural pathway that connects the substantia nigra with the striatum). These marks are detectable using neuroimaging techniques, which have become an ordinary practice in the diagnosis of neurodegenerative disorders as PD. The deficiency of striatal dopamine can be assessed using a variety of nuclear medicine techniques targeted to dopamine transporters (DaT). For example, the ¹²³I-ioupane (also known by its tradename DaTSCAN) is a widely-used radioligand that binds to the dopamine transporters in the striatum and allows visualizing the presynaptic striatal dopamine deficiency state with high sensitivity. This drug has been successfully used to differentiate PD from healthy controls (Booij et al., 1998; Winogrodzka et al., 2003; Towey et al., 2011; Illán et al., 2012; Segovia et al., 2012). However, the effectiveness of DaTSCAN to distinguish PD from atypical parkinsonian syndromes (APS), such as multiple system atrophy (MSA; Wenning et al., 2013) and progressive supranuclear palsy

(PSP; Williams and Lees, 2009), is limited due to DaT is reduced in both cases. Alternative radioligands have been proposed for this problem. For example, ^{18}F -Fludeoxyglucose (^{18}F -FDG) PET data, which allow evaluating the glucose metabolism of the brain, were successfully used to discriminate APS from PD in Garraux et al. (2013) and Ghaemi et al. (2002). A postsynaptic striatal deficiency can be also demonstrated by means of specific radioligands that focus on $D_{2/3}$ striatal dopamine receptors. ^{123}I -Iodobenzamide (^{123}I -IBZM) SPECT, ^{11}C -Raclopride PET and ^{18}F -Desmethoxyfallypride (^{18}F -DMFP) PET have suggested for this purpose (Cordes et al., 1991; Antonini et al., 1997; Stark et al., 2007; Vernaleken et al., 2007). In addition, the latter was recently used to separate idiopathic and non-idiopathic parkinsonian syndromes with high accuracy (la Fougère et al., 2010).

Nowadays, the visual examination of the neuroimages is combined with modern computer systems that automatically analyze the data and are able to estimate their class (pathological or not). In the case of PD diagnosis, the classical approach addresses the problem directly, i.e., by quantifying the loss of striatal dopamine neurons (Antonini et al., 1997; Constantinescu et al., 2011; Garraux et al., 2013). However, modern computer systems examine the images, looking for the patterns that characterize the studied disease. Two approaches have been proposed: univariate and multivariate. On the one hand, univariate methods analyze each voxel independently, without considering the interactions between voxels. The most relevant univariate analysis approach to date is the widely used Statistical Parametric Mapping (SPM) (Friston et al., 2007). This approach has been used in a number of neuroimaging studies, including the diagnosis of neurological disorders (Hosaka et al., 2002; Wang et al., 2007; Perani et al., 2014). On the other hand, multivariate approaches consider all the voxel as a whole, evaluating the correlation of activation across brain regions (Habeck, 2010; Schrouff et al., 2013). A substantial part of multivariate methods are based on machine learning algorithms such as decision trees (Quinlan, 1986) or support vector machine (SVM; Vapnik, 1999). They have been successfully used to assist the diagnosis of Parkinsonism using different neuroimage modalities (Cordes et al., 1991; Dodel et al., 2003; Illán et al., 2012; Segovia et al., 2012; Nair et al., 2013; Mudali et al., 2015). Bayesian approaches (Ben-Gal, 2008) were also used for this purpose. In Towey et al. (2011), a naive Bayes classifier is used along with Principal Component Analysis in order to separate parkinsonian and non-parkinsonian syndromes. Several Bayesian methods were also evaluated in Morales et al. (2013), achieving accuracy rates over 90% when predicting dementia development in PD patients.

In this work, we demonstrate a method based on SVM classification and Bayesian networks to separate idiopathic PD from APS using ^{18}F -DMFP PET data. Initially we performed a selection of the regions of interest (ROIs) for this problem using a t -test. The selected regions were then evaluated by means of a SVM classifier. Subsequently, a Bayesian network was developed using the outputs of the SVM classification procedure. Finally, the Bayesian network was used to classify new unseen data. This method was evaluated using a database with 87 DMFP neuroimages labeled as idiopathic PD, MSA or PSP. Accuracy rates over 78% were obtained when separating

PD from APS (MSA or PSP), outperforming other previous approaches.

2. MATERIALS AND METHODS

2.1. Neuroimaging Data

Neuroimaging data used in this work was collected in a longitudinal study carried out in the University of Munich (la Fougère et al., 2010). Eighty-seven (87) patients with parkinsonism, previously confirmed by a ^{123}I -FP-CIT SPECT scan according to widely accepted criteria (Koch et al., 2005), were undergone $D_{2/3}$ receptor imaging with ^{18}F -DMFP. Sixty (60) min after the radiopharmaceutical injection the neuroimaging data were acquired by means of a ECAT EXACT HR⁺ PET scanner (Siemens/CTI). The emission recording consisted of 3 frames of 10 min each, acquired in 3-dimensional mode. The resulting images were reconstructed as 128×128 matrices of 2×2 mm voxels by filtered backprojection using a Hann filter with a cutoff frequency of 0.5 Nyquist and corrected for randoms, dead time, and scatter.

The patients were clinically monitored during the following years. Two years after the data acquisition, the neuroimages were labeled by experienced clinicians on the basis of last observations. According to the United Kingdom Parkinson Disease Society Brain Bank Diagnostic Criteria for Parkinson Disease (Hughes et al., 2002), the second consensus statement on the diagnosis of multiple-system atrophy (Gilman et al., 2008) and the established criteria for the diagnosis of progressive supranuclear palsy (Litvan et al., 1996), 3 groups were defined: idiopathic PD, MSA, and PSP patients. Demographic details of the groups are shown in **Table 1**.

In order to ensure that any given voxel in different images refers to the same anatomical position across the brains, all the images were spatially normalized using the template matching approach implemented in the SPM software (version 8). An *ad-hoc* template was generated for this procedure: First, idiopathic PD neuroimages were registered to a randomly chosen one. The resulting images and their hemisphere midplane reflections were then averaged, obtaining a symmetric image. That way, effects due patients affected by non-bilateral PD variants are minimized (Djaldetti et al., 2006). Finally, this image was smoothed and used as template.

After the spatial normalization, the intensity of the images was also normalized to a value I_{max} , obtained by averaging the 0.1%

TABLE 1 | Sex and age information of the groups of patients considered in this work.

	#	Sex		Age		
		M	F	μ	σ	Range
Idiopathic PD	39	22	17	61.38	11.14	35-81
MSA	24	20	4	68.42	10.73	43-85
PSP	24	12	12	69.29	7.33	55-84

μ and σ stand for the average and the standard deviation respectively.

of the highest intensities per image, as described in Saxena et al. (1998).

2.2. Background on Bayesian Networks

Bayesian networks (a.k.a. belief networks) are statistical models belonging to the family of probabilistic graphical models that combine principles from graph theory, probability theory, computer science, and statistics (Ben-Gal, 2008). In particular, they are directed acyclic graphs (DAG) in which nodes represent random variables in the Bayesian sense (observable quantities, latent variables, unknown parameters or hypotheses) and edges between the nodes represent probabilistic dependencies among the corresponding random variables. A Bayesian network defines a unique joint probability distribution given by:

$$p(x^{(1)}, \dots, x^{(M)}|y) = \prod_{m=1}^M p(x^{(m)}|Pa_{x^{(m)}}) = \prod_{m=1}^M \theta_{x^{(m)}|Pa_{x^{(m)}}} \quad (1)$$

where $\mathbf{x} = \{x^{(1)}, \dots, x^{(M)}\}$ and y respectively denote a set of random variables and a property of them. In our case \mathbf{x} represents the neuroimaging features for a given patient and y is the label assigned to that patient. θ is the set of parameters that quantifies the conditional probability distribution in the network and $Pa_{x^{(m)}}$ represents the set of parents of $x^{(m)}$ in the graph.

When the topology and/or the parameters of the network are unknown, they can be estimated using a set of training data. Two approaches are commonly used to learn the network structure: constraint-based and search-and-score. The former starts with a fully connected graph, and remove edges if certain conditions are satisfied in the training data. The latter approach performs an exhaustive search in the space of all possible structures, which are evaluated using a predefined scoring function.

Once the structure is learned and the parameters are fixed, the network can be used for inference. In this case, the Bayesian network encodes a disruption $p(x^{(1)}, \dots, x^{(M)}, y)$ so that, given a set of features, $x^{(1)}, \dots, x^{(M)}$, it returns the label y that maximizes the posterior probability $P(y|x^{(1)}, \dots, x^{(M)})$, which is trivially derived from Equation 1 using the definition of conditional probability and the chain rule.

2.3. Background on Support Vector Machines

Support vector machine is a supervised learning method derived from the statistical learning theory, which was developed by Vladimir Vapnik in late 90s (Vapnik, 1999). A SVM classifier builds a function $f: \mathbb{R}^D \rightarrow \{\pm 1\}$ using the training data (n D -dimensional patterns, $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, and their class labels, $\{y_1, \dots, y_n\}$) so that f is able to predict the label y_i of a new example \mathbf{x}_i .

SVM can use kernelized inputs that allow us learning a nonlinear function or decision boundary. More sophisticated variants based on multiple kernel learning (MKL) define the kernel function as the combination of other simpler kernels (Gonen and Alpaydin, 2011):

$$k(\mathbf{x}_i \mathbf{x}_j) = f(\{k_p(\mathbf{x}_i^p, \mathbf{x}_j^p)\}_{p=1}^P) \quad (2)$$

where kernel k is computed as the combination of P kernels each of which taking a feature representation of data instances: $\mathbf{x}_i = \{\mathbf{x}_i^p\}_{p=1}^P$ where $\mathbf{x}_i^p \in \mathbb{R}^{D_p}$ and D_p is the dimensionality of the corresponding feature representation.

2.4. Selection of Regions of Interest

¹⁸F-DMFP has a high binding affinity for dopamine transporters in the striatal region of the brain. For this reason, neuroimaging studies based on it usually focuses on that region. However, ¹⁸F-DMFP PET neuroimages contain a substantial part of the total intensity in regions other than the striatum. In order to reveal the most important regions to separate idiopathic PD from APS we carried out an univariate analysis. Specifically we performed a two-sample t -test comparing both populations under the hypothesis of data corresponding to idiopathic PD patients have lower intensity than those from APS patients. The test was carried out using the SPM software (version 8) and a smoothed version of the neuroimaging data. The full-width at half maximum of the Gaussian smoothing kernel was fixed at 8 mm. The resulting map, thresholded at $p < 0.001$ (uncorrected), is shown in **Figure 1**.

Finally, the t -test map was matched with the Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002) and the regions of the atlas with the highest proportion of highlighted voxels in the t -test map were selected.

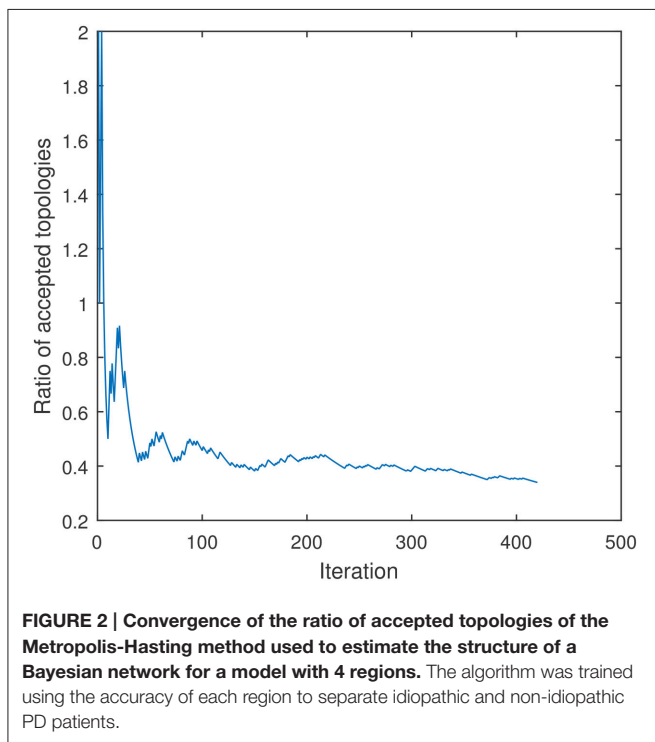
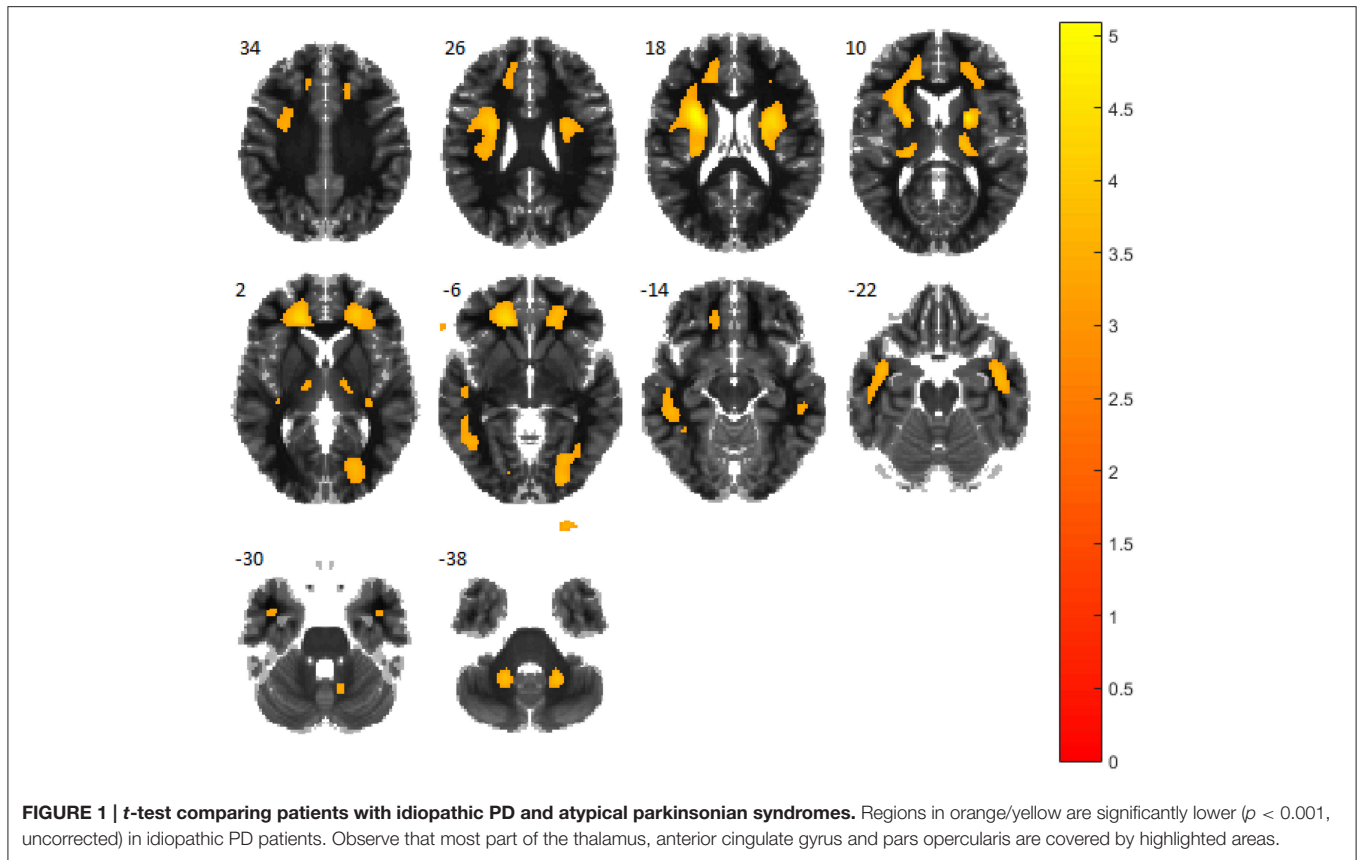
2.5. Multivariate Analysis Based on Support Vector Machines and Bayesian Networks

Once a reduced set of regions of interest has been selected, the accuracy of each region to discriminate between groups was computed using a SVM classifier and a leave-one-out (LOO) cross-validation (CV) scheme. As a result, an accuracy value was assigned to each region.

The structure of the Bayesian network is then estimated by means of a Metropolis-Hastings (MH) algorithm (Metropolis et al., 1953; Hastings, 1970), a Markov chain Monte Carlo (MCMC) method that converges after about 200 iterations (see **Figure 2**) and is more efficient than the more popular search-and-score approach. This algorithm was implemented to use the Bayesian information criterion as the score function to find the optimal structure (Heckerman et al., 1995).

The parameters of the model were calculated by means of a maximum a posteriori scheme, which assumes that all the variables are fully observed. The accuracy of each region previously computed was used in this procedure as training data. **Figure 3** shows a Bayesian network for a model with 4 regions. Observe that the labels are included in the graph as an additional node. This allows using the network for inference (Illan et al., 2014).

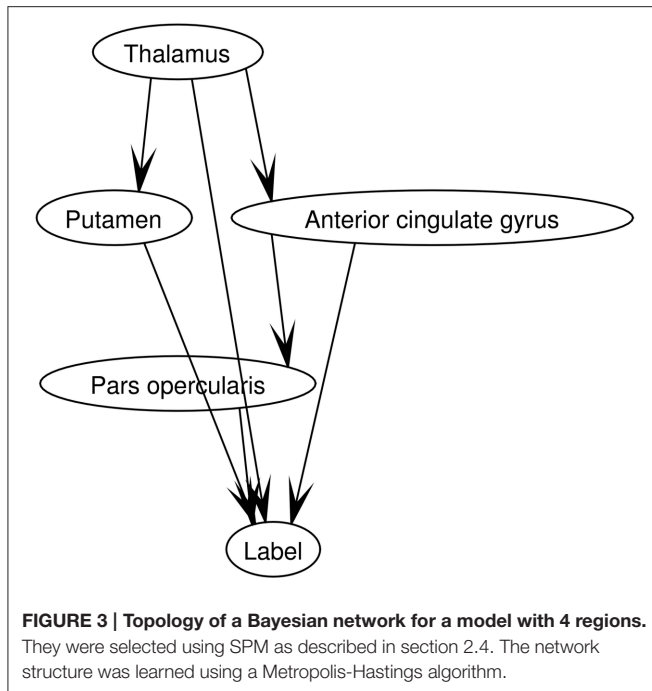
Finally, the Bayesian network was used to classify new unseen data by selecting the label that maximizes the posterior probability. If the two possible labels have equal probability (0.5 in both cases), the label with the most *votes* was selected.



3. EXPERIMENTS AND RESULTS

The methodology proposed above was evaluated using the neuroimaging data described in section 2.1. In order to avoid biased results, a LOO-CV scheme was implemented. Since another LOO-CV loop was used to estimate the accuracy of the regions, this resulted in a 2 levels (nested) CV. This strategy has been suggested as an effective method to evaluate the risk of a classification procedure (Varma and Simon, 2006), especially when the database is small. The Bayes Net Toolbox (Murphy, 2001) and LIBSVM (Chang and Lin, 2011) package were respectively used for structure learning and SVM classification. They were parametrized to use the default values, including the cost parameter of the classifier, C , which was fixed to the commonly accepted value of 1. The SVM classifier used linear kernels. The pseudo-code that describes all the experiments is shown in Algorithm 1.

Table 2 shows the accuracy measures obtained by the proposed method. The results are compared with those obtained by other methods based on SVM classification: (i) a *Voxel-As-Feature* approach, which consists on using all the voxels as feature, (ii) the classical approach for PD diagnosis consisting on using only voxels in the striatum area, (iii) a method that uses the voxels belonging to the selected regions as a whole, (iv) a multiple classifier approach with a classifier per selected region



and a majority voting strategy, and (v) a multiple kernel learning algorithm with a kernel per region.

4. DISCUSSION AND CONCLUSIONS

Distinguishing between Parkinson's disease and atypical parkinsonian syndromes is still a challenge due to both disorders have similar symptoms (Litvan, 1999). Still, the methodology we demonstrated achieved an accuracy rate over 78% and a good trade-off between sensitivity and specificity. These results suggest the proposed method is suitable to assist the diagnosis of PD and confirm the usefulness of DMFP data for this purpose.

Comparison with previous works in terms of classification accuracy is difficult because of the lack of studies using DMFP data for the same purpose. The separation of PD from other non-idiopathic parkinsonian syndromes have been previously addressed using multiclass classification procedures and MRI or PET data with ^{18}F -FDOPA, ^{18}F -FDG or ^{11}C -RACLO (Ghaemi et al., 2002; Eckert et al., 2004, 2005; Spetsieris et al., 2009; Garraux et al., 2013). Reported accuracy rates vary from 68.5% up to 90%. This value depends on the database (the number of groups, the criteria to assign the labels, etc.) and the methodology (usually a binary classification is simpler than a multiclass procedure). The methodology proposed in this work achieved an accuracy rate of 78.16% and outperformed other previous approaches in a fear comparison (using the same data and classification approach), as shown in **Table 2**. The relatively low accuracy rates obtained, in general, by all the methods evaluated in this work is explained by the neuroimages we used, which correspond to early stages of the disease. Observe that the neuroimaging data were acquired

Algorithm 1: Evaluation procedure

Input: The image database, $D = I_1, \dots, I_n$, and the labels, $Y = Y_1, \dots, Y_n$

Output: Accuracy measures of the method

```

foreach image  $i \in D$  do
   $D_T =$  All the images in  $D$  but  $i$ 
   $Y_T =$  Labels for images in  $D_T$ 
   $m =$  Compute t-test map using data in  $D_T$  and the
  groups defined by  $Y_T$ 
   $R =$  Select ROIs from  $m$ 
  foreach region  $r \in R$  do
    foreach image  $j \in D_T$  do
       $D_{T_r} =$  All the images in  $D_T$  but  $j$ 
       $Y_{T_r} =$  Labels for images in  $D_{T_r}$ 
       $C =$  Train a SVM classifier using  $D_{T_r}$  and  $Y_{T_r}$ 
       $S_{r,j} =$  Test image  $j$  in  $C$ 
    end
     $S_r =$  Compute accuracy for region  $r$  using  $S_{r,j}$ 
  end
  // Create network
   $G =$  Learn graph topology using  $S_r$ 
   $\theta =$  Compute conditional probability distributions
   $B =$  Create Bayesian network using  $G$  and  $\theta$ 
  // Inference
  foreach region  $r \in R$  do
     $C =$  Train a SVM classifier using voxels at region  $r$ 
    from images in  $D_T$  and  $Y_T$ 
     $E_r =$  Test voxels at region  $r$  from image  $i$  in  $C$ 
  end
  foreach label  $l \in \{-1, 1\}$  do
     $P_l =$  Compute posterior probability using  $B$  and  $E$ 
  end
  if  $P_{-1} = P_1$  then
     $\hat{Y}_i = \text{sign}(\sum E_r)$  // Majority voting
  else if  $P_{-1} > P_1$  then
     $\hat{Y}_i = -1$ 
  else
     $\hat{Y}_i = 1$ 
  end
end
  Estimate the accuracy measures comparing  $\hat{Y}$  and  $Y$ 
  
```

during the first visit, 2 years before assigning the final diagnosis that was used to label the data. Furthermore, it is worth noting that the database was clinically labeled, what introduced an error due to the intrinsic limitations of the clinical assessment (Jobst et al., 1998), and the generalization of the classification procedure should be interpreted from this perspective.

In addition to the putamen, whose relation with PD is widely accepted, other regions were identified by the univariate analysis. Namely, the thalamus, anterior cingulate gyrus and pars

TABLE 2 | Classification performance of the proposed algorithm compared with other approaches.

	Accuracy %	Sensitivity %	Specificity %	Positive likelihood	Negative likelihood
All the voxel in the brain	70.11	61.54	77.08	2.69	0.50
Only striatum	73.56	69.23	77.08	3.02	0.40
Selected regions as a whole	70.11	66.67	72.92	2.46	0.46
Multiple SVM (majority voting)	74.71	74.36	75.00	2.97	0.34
Multiple kernel SVM	75.86	71.79	79.17	3.45	0.36
Proposed method (Bayesian network)	78.16	76.92	79.17	3.69	0.29

opercularis (a subregion of the inferior frontal gyrus) showed differences when comparing PD and APS patients (see **Figure 1**). These three areas has been previously reported as affected by PSP and not affected by MSA or PD what facilitates separating PSP patients from the PD group (Zgaljardic and Feigin, 2004; Varrone et al., 2007; Messina et al., 2011).

One of the main feature of the proposed method is the combination of univariate and multivariate analyses. Both approaches are widely used for diagnosis purposes, however, the former is more suitable for groups comparison while the latter can be easily used for classification. For that reason, we used a univariate analysis for the selection of the regions of interest and a multivariate analysis to separate the groups. This second analysis was based on SVM classification and Bayesian networks. The data was first analyzed, region by region, in a SVM classifier and its outputs were then managed by a Bayesian network to provide a final output. Thus, the Bayesian network can be seen as a way of weighting the SVM decisions for the individual regions. Therefore, this methodology can be used as an alternative to the majority voting strategy and other approaches to deal with multiple classifier decisions in systems for a single classification problem. The main advantage of Bayesian networks in this context is their ability to take into account the information about the relations between regions. The obtained results suggest that this information is useful to separate PD and APS patients. If reproduced by other studies, it could improve the diagnosis of these disorders.

As other computer aided diagnosis systems, the method proposed in this work could be a valuable tool for the clinical practice. Using a database properly labeled, the system is able to analyze neuroimaging data from a new patient and estimate the disorder he/she suffers. However, this procedure should be supervised by experienced clinicians to corroborate that the diagnosis estimated by the system is consistent with the patient's symptoms.

ACKNOWLEDGMENTS

The authors are grateful to Madeleine Schuberth and Matthias Brendel for their help in data management. This work is part of a project approved by the Andalucía Talent Hub Program launched by the Andalusian Knowledge Agency, co-funded by the European Union's Seventh Framework Program, Marie Skłodowska-Curie actions (COFUND Grant Agreement no 291780) and the Ministry of Economy, Innovation, Science and Employment of the Junta de Andalucía. The work was also supported by the University of Granada (Spain), the University for Munich (Germany), the MICINN (Spain) under the TEC2012-34306 project and the Consejera de Innovación, Ciencia y Empresa (Junta de Andalucía, Spain) under the P11-TIC-7103 excellence project. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

- Antonini, A., Leenders, K. L., Vontobel, P., Maguire, R. P., Missimer, J., Psylla, M., et al. (1997). Complementary PET studies of striatal neuronal function in the differential diagnosis between multiple system atrophy and parkinson's disease. *Brain* 120, 2187–2195. doi: 10.1093/brain/120.12.2187
- Ben-Gal, I. (2008). "Bayesian networks," in *Encyclopedia of Statistics in Quality and Reliability* (Hoboken, NJ: John Wiley & Sons, Ltd).
- Booij, J., Habraken, J. B., Bergmans, P., Tissingh, G., Winogrodzka, A., Wolters, E. C., et al. (1998). Imaging of dopamine transporters with iodine-123-FP-CIT SPECT in healthy controls and patients with Parkinson's disease. *J. Nuclear Med.* 39, 1879–1884.
- Chang, C., and Lin, C. (2011). LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2, 27:1–27:27. doi: 10.1145/1961189.1961199
- Constantinescu, C. C., Coleman, R. A., Pan, M., and Mukherjee, J. (2011). Striatal and extrastriatal microPET imaging of D2/D3 dopamine receptors in rat brain with [18F]Fallypride and [18F]Desmethoxyfallypride. *Synapse* 65, 778–787. doi: 10.1002/syn.20904
- Cordes, M., Henkes, H., Laudahn, D., Bräu, H., Kramp, W., Girke, W., et al. (1991). Initial experience with SPECT examinations using [123I]IBZM as a d2-dopamine receptor antagonist in parkinson's disease. *Eur. J. Radiol.* 12, 182–186. doi: 10.1016/0720-048X(91)90069-8
- Djalldetti, R., Ziv, I., and Melamed, E. (2006). The mystery of motor asymmetry in Parkinson's disease. *Lancet Neurol.* 5, 796–802. doi: 10.1016/S1474-4422(06)70549-X
- Dodel, R. C., Höffken, H., Möller, J. C., Bornschein, B., Klockgether, T., Behr, T., et al. (2003). Dopamine transporter imaging and SPECT in diagnostic work-up of parkinson's disease: a decision-analytic approach. *Mov. Disord.* 18(Suppl. 7), S52–S62. doi: 10.1002/mds.10580
- Eckert, T., Barnes, A., Dhawan, V., Frucht, S., Gordon, M. F., Feigin, A. S., et al. (2005). FDG PET in the differential diagnosis of parkinsonian disorders. *Neuroimage* 26, 912–921. doi: 10.1016/j.neuroimage.2005.03.012

- Eckert, T., Sailer, M., Kaufmann, J., Schrader, C., Peschel, T., Bodammer, N., et al. (2004). Differentiation of idiopathic parkinson's disease, multiple system atrophy, progressive supranuclear palsy, and healthy controls using magnetization transfer imaging. *Neuroimage* 21, 229–235. doi: 10.1016/j.neuroimage.2003.08.028
- Friston, K. J., Ashburner, J. T., Kiebel, S. J., Nichols, T. E., and Penny, W. D. (eds.). (2007). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Amsterdam: Academic Press.
- Garraux, G., Phillips, C., Schrouff, J., Kreisler, A., Lemaire, C., Degueldre, C., et al. (2013). Multiclass classification of FDG PET scans for the distinction between parkinson's disease and atypical parkinsonian syndromes. *Neuroimage* 2, 883–893. doi: 10.1016/j.nicl.2013.06.004
- Ghaemi, M., Hilker, R., Rudolf, J., Sobesky, J., and Heiss, W. (2002). Differentiating multiple system atrophy from parkinson's disease: contribution of striatal and midbrain MRI volumetry and multi-tracer PET imaging. *J. Neurol. Neurosurg. Psychiatr.* 73, 517–523. doi: 10.1136/jnnp.73.5.517
- Gilman, S., Wenning, G. K., Low, P. A., Brooks, D. J., Mathias, C. J., Trojanowski, J. Q., et al. (2008). Second consensus statement on the diagnosis of multiple system atrophy. *Neurology* 71, 670–676. doi: 10.1212/01.wnl.0000324625.00404.15
- Gonen, M., and Alpaydin, E. (2011). Multiple kernel learning algorithms. *J. Mach. Learn. Res.* 12, 2211–2268.
- Greenberg, D., Aminoff, M., and Simon, R. (2012). *Clinical Neurology*. 8th Edn. New York, NY: McGraw-Hill Professional.
- Habeck, C. G. (2010). Basics of multivariate analysis in neuroimaging data. *J. Vis. Exp.* 41:e1988. doi: 10.3791/1988
- Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika* 57, 97–109. doi: 10.1093/biomet/57.1.97
- Heckerman, D., Geiger, D., and Chickering, D. M. (1995). Learning bayesian networks: the combination of knowledge and statistical data. *Mach. Learn.* 20, 197–243. doi: 10.1007/BF00994016
- Hosaka, K., Ishii, K., Sakamoto, S., Mori, T., Sasaki, M., Hirono, N., et al. (2002). Voxel-based comparison of regional cerebral glucose metabolism between PSP and corticobasal degeneration. *J. Neurol. Sci.* 199, 67–71. doi: 10.1016/S0022-510X(02)00102-8
- Hughes, A. J., Daniel, S. E., Ben-Shlomo, Y., and Lees, A. J. (2002). The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. *Brain* 125, 861–870. doi: 10.1093/brain/awf080
- Illan, I. A., Górriz, J. M., Ramírez, J., and Meyer-Base, A. (2014). Spatial component analysis of MRI data for alzheimer's disease diagnosis: a bayesian network approach. *Front. Comput. Neurosci.* 8:156. doi: 10.3389/fncom.2014.00156
- Illán, I. A., Górriz, J. M., Ramírez, J., Segovia, F., Jiménez-Hoyuela, J. M., and Lozano, S. J. O. (2012). Automatic assistance to parkinson's disease diagnosis in DaTSCAN SPECT imaging. *Med. Phys.* 39, 5971–5980. doi: 10.1118/1.4742055
- Jobst, K. A., Barnetson, L. P., and Shepstone, B. J. (1998). Accurate prediction of histologically confirmed alzheimer's disease and the differential diagnosis of dementia: the use of NINCDS-ADRDA and DSM-III-R criteria, SPECT, x-ray CT, and apo e4 in medial temporal lobe dementias. oxford project to investigate memory and aging. *Int. Psychogeriatr.* 10, 271–302. doi: 10.1017/S1041610298005389
- Koch, W., Radau, P. E., Hamann, C., and Tatsch, K. (2005). Clinical testing of an optimized software solution for an automated, observer-independent evaluation of dopamine transporter SPECT studies. *J. Nuclear Med.* 46, 1109–1118.
- la Fougère, C., Pöpperl, G., Levin, J., Wängler, B., Böning, G., Uebles, C., et al. (2010). The value of the dopamine d2/3 receptor ligand 18F-Desmethoxyfallypride for the differentiation of idiopathic and nonidiopathic parkinsonian syndromes. *J. Nuclear Med.* 51, 581–587. doi: 10.2967/jnumed.109.071811
- Litvan, I. (1999). Recent advances in atypical parkinsonian disorders. *Curr. Opin. Neurol.* 12, 441–446. doi: 10.1097/00019052-199908000-00011
- Litvan, I., Agid, Y., Calne, D., Campbell, G., Dubois, B., Duvoisin, R. C., et al. (1996). Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDS-SPSP international workshop. *Neurology* 47, 1–9. doi: 10.1212/WNL.47.1.1
- Messina, D., Cerasa, A., Condino, F., Arabia, G., Novellino, F., Nicoletti, G., et al. (2011). Patterns of brain atrophy in parkinson's disease, progressive supranuclear palsy and multiple system atrophy. *Parkins. Rel. Disord.* 17, 172–176. doi: 10.1016/j.parkreldis.2010.12.010
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21, 1087–1092. doi: 10.1063/1.1699114
- Morales, D. A., Vives-Gilbert, Y., Gómez-Ansón, B., Bengoetxea, E., Larrañaga, P., Bielza, C., et al. (2013). Predicting dementia development in parkinson's disease using bayesian network classifiers. *Psychiat. Res.* 213, 92–98. doi: 10.1016/j.psychres.2012.06.001
- Mudali, D., Teune, L. K., Renken, R. J., Leenders, K. L., and Roerdink, J. B. T. M. (2015). Classification of parkinsonian syndromes from FDG-PET brain data using decision trees with SSM/PCA features. *Comput. Math. Methods Med.* 2015:e136921. doi: 10.1155/2015/136921
- Murphy, K. P. (2001). The bayes net toolbox for MATLAB. *Comput. Sci. Stat.* 33, 1024–1034.
- Nair, S. R., Tan, L. K., Mohd Ramli, N., Lim, S. Y., Rahmat, K., and Mohd Nor, H. (2013). A decision tree for differentiating multiple system atrophy from parkinson's disease using 3-T MR imaging. *Eur. Radiol.* 23, 1459–1466. doi: 10.1007/s00330-012-2759-9
- Perani, D., Della Rosa, P. A., Cerami, C., Gallivanone, F., Fallanca, F., Vanoli, E. G., et al. (2014). Validation of an optimized SPM procedure for FDG-PET in dementia diagnosis in a clinical setting. *Neuroimage* 6, 445–454. doi: 10.1016/j.nicl.2014.10.009
- Quinlan, J. R. (1986). Induction of decision trees. *Mach. Learn.* 1, 81–106. doi: 10.1007/BF00116251
- Saxena, P., Pavel, D. G., Quintana, J. C., and Horwitz, B. (1998). "An automatic thresholdbased scaling method for enhancing the usefulness of Tc-HMPAO SPECT in the diagnosis of Alzheimers disease," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI, Lecture Notes in Computer Science*, Vol. 1496, eds W. M. Wells, A. Colchester, and S. Delp (Cambridge, MA: Springer Berlin Heidelberg), 623–630. doi: 10.1007/BFb0056248
- Schrouff, J., Rosa, M. J., Rondina, J. M., Marquand, A. F., Chu, C., Ashburner, J., et al. (2013). PRoNTto: pattern recognition for neuroimaging toolbox. *Neuroinformatics* 11, 319–337. doi: 10.1007/s12021-013-9178-1
- Segovia, F., Górriz, J. M., Ramírez, J., Álvarez, I., Jiménez-Hoyuela, J. M., and Ortega, S. J. (2012). Improved parkinsonism diagnosis using a partial least squares based approach. *Med. Phys.* 39, 4395–4403. doi: 10.1118/1.4730289
- Spetsieris, P. G., Ma, Y., Dhawan, V., and Eidelberg, D. (2009). Differential diagnosis of parkinsonian syndromes using PCA-based functional imaging features. *Neuroimage* 45, 1241–1252. doi: 10.1016/j.neuroimage.2008.12.063
- Stark, D., Piel, M., Hübner, H., Gmeiner, P., Gründer, G., and Rösch, F. (2007). *In vitro* affinities of various halogenated benzamide derivatives as potential radioligands for non-invasive quantification of d2-like dopamine receptors. *Bioorgan. Med. Chem.* 15, 6819–6829. doi: 10.1016/j.bmc.2007.07.017
- Towey, D. J., Bain, P. G., and Nijran, K. S. (2011). Automatic classification of 123I-FP-CIT (DaTSCAN) SPECT images. *Nuclear Med. Commun.* 32, 699–707. doi: 10.1097/MNM.0b013e328347cd09
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI Single-Subject brain. *Neuroimage* 15, 273–289. doi: 10.1006/nimg.2001.0978
- Vapnik, V. (1999). *The Nature of Statistical Learning Theory*. 2nd Edn. New York, NY: Springer.
- Varma, S., and Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* 7:91. doi: 10.1186/1471-2105-7-91
- Varrone, A., Pagani, M., Salvatore, E., Salmaso, D., Sansone, V., Amboni, M., Nobili, F., et al. (2007). Identification by [99mTc]ECD SPECT of anterior cingulate hypoperfusion in progressive supranuclear palsy, in comparison with parkinson's disease. *Eur. J. Nuclear Med. Mol. Imaging* 34, 1071–1081. doi: 10.1007/s00259-006-0344-7
- Vernaleken, I., Weibrich, C., Siessmeier, T., Buchholz, H., Rösch, F., Heinz, A., et al. (2007). Asymmetry in dopamine d2/3 receptors of caudate nucleus is lost with age. *Neuroimage* 34, 870–878. doi: 10.1016/j.neuroimage.2006.10.013

- Wang, D. J., Zuo, C. T., Jiang, Y. P., Guan, Y. H., Chen, Z. P., Xiang, J. D., et al. (2007). 18F-FP-CIT PET imaging and SPM analysis of dopamine transporters in parkinson's disease in various hoehn & yahr stages. *J. Neurol.* 254, 185–190. doi: 10.1007/s00415-006-0322-9
- Wenning, G. K., Geser, F., Krismer, F., Seppi, K., Duerr, S., Boesch, S., et al. (2013). The natural history of multiple system atrophy: a prospective european cohort study. *Lancet Neurol.* 12, 264–274. doi: 10.1016/S1474-4422(12)70327-7
- Williams, D. R., and Lees, A. J. (2009). Progressive supranuclear palsy: clinicopathological concepts and diagnostic challenges. *Lancet Neurol.* 8, 270–279. doi: 10.1016/S1474-4422(09)70042-0
- Winogrodzka, A., Bergmans, P., Booij, J., van Royen, E. A., Stoof, J. C., and Wolters, E. C. (2003). [123I]beta-CIT SPECT is a useful method for monitoring dopaminergic degeneration in early stage Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* 74, 294–298. doi: 10.1136/jnnp.74.3.294
- Zgaljardic, D. J., and Feigin, A. (2004). Neuroimaging of parkinson's disease and atypical parkinsonism. *Curr. Neurol. Neurosci. Rep.* 4, 284–289. doi: 10.1007/s11910-004-0053-1

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Segovia, Illán, Górriz, Ramírez, Rominger and Levin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.