Multi-class parkinsonian disorders classification with quantitative MR markers and graph-based features using support vector machines

Rita Morisi a,1, David Neil Manners b, c,1, Giorgio Gnecco d, Nico Lanconelli d, Claudia Testa b, c, Stefania Evangelisti b, c, Lia Talozzi b, c, Laura Ludovica Gramegna b, c, Claudio Bianchini b, c, Giovanna Calandra-Buonaura c, e, Luisa Sambati f, Giulia Giannini c, Pietro Cortelli c, e, Caterina Tonon b, c,2, Raffaele Lodi b, c,1,2

a IMT School for Advanced Studies, Piazza S. Ponziano, 6, 55100, Lucca, Italy
b Functional MR Unit, Policlinico S. Orsola - Malpighi, Via Massarenti 9, 40138, Bologna, Italy
c Department of Biomedical and NeuroMotor Sciences, University of Bologna, Via U. Foscolo 7, 40123, Bologna, Italy
d Department of Physics and Astronomy, University of Bologna, Viale Berti-Pichat 6/2, 40127, Bologna, Italy
e IRCCS Institute of Neurological Sciences of Bologna, Via Altura 3, 40139, Bologna, Italy
f Corresponding author. Functional MR Unit, Policlinico S. Orsola – Malpighi, Department of Biomedical and NeuroMotor Sciences (DiBiNeM), University of Bologna, Via Massarenti 9, 40138, Bologna, Italy.
E-mail address: raffaele.lodi@unibo.it (R. Lodi).
1 Contributed equally.
2 Contributed equally.

A R T I C L E   I N F O

Article history:
Received 3 August 2017
Received in revised form 30 October 2017
Accepted 27 November 2017

Keywords:
Parkinsonian disorders
MR markers
Feature selection
Support vector machines
Graph-based features

A B S T R A C T

Background and purpose: In this study we attempt to automatically classify individual patients with different parkinsonian disorders, making use of pattern recognition techniques to distinguish among several forms of parkinsonisms (multi-class classification), based on a set of binary classifiers that discriminate each disorder from all others.

Methods: We combine diffusion tensor imaging, proton spectroscopy and morphometric-volumetric data to obtain MR quantitative markers, which are provided to support vector machines with the aim of recognizing the different parkinsonian disorders. Feature selection is used to find the most important features for classification. We also exploit a graph-based technique on the set of quantitative markers to extract additional features from the dataset, and increase classification accuracy.

Results: When graph-based features are not used, the MR markers that are most frequently automatically extracted by the feature selection procedure reflect alterations in brain regions that are also usually considered to discriminate parkinsonisms in routine clinical practice. Graph-derived features typically increase the diagnostic accuracy, and reduce the number of features required.

Conclusions: The results obtained in the work demonstrate that support vector machines applied to multimodal brain MR imaging and using graph-based features represent a novel and highly accurate approach to discriminate parkinsonisms, and a useful tool to assist the diagnosis.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Degenerative parkinsonisms, such as Idiopathic Parkinson’s Disease (PD), Progressive Supranuclear Palsy (PSP) and Multiple System Atrophy (MSA), with the cerebellar (MSA-C) and parkinsonian (MSA-P) variants, are chronic progressive diseases characterized primarily by movement impairment accompanied by various autonomic, cognitive and behavioral alterations [1]. Although these disorders are characterized by different clinical features, response to pharmacological treatment and prognosis, in vivo differential diagnosis is often challenging because of the possible clinical overlap, and a definite diagnosis can be achieved only post mortem [1]. Various in vivo biomarkers have been identified that may improve the accuracy of clinical diagnostic criteria [2]. In particular, advanced MR quantitative markers of brain microstructure, metabolism and regional atrophy obtained from Diffusion Tensor Imaging (DTI), proton spectroscopy (1H-MRS) and morphometric-volumetric analysis respectively, have demonstrated high accuracy in differentiating subjects with Parkinson’s...
disease, atypical parkinsonism and essential tremor [3–10].

An approach combining these quantitative MR markers may also improve automatic diagnostic classification at an individual level [10], improving differential diagnosis when combined with neuroradiological expertise.

Supervised learning classifies unknown cases by comparing features of such cases with those of known cases. Among supervised learning techniques, Support Vector Machines (SVMs) combine sets of discriminative variables to form a large dimensional feature vector, then exploit the training data to perform a linear classification in the feature space [11].

Recent studies have evaluated the usefulness of such an approach applied to structural brain MR markers to discriminate patients with PD from other groups including those with PSP [12], healthy controls [13] and atypical parkinsonisms [14,15]. Although most studies focused on binary problems concerning the differentiation of PD from other movement disorders, another study used a multi-class classification approach to distinguish different forms of parkinsonism (PD, PSP, MSA-C, and MSA-P) [16] using morphometric features. We have previously used SVMs to study both classification approaches [17].

Graph theory can also provide useful tools to model complex data, analyzing datasets as nodes (here, study participants) linked by edges (MR features), whose similarity may be evaluated using appropriate metrics [18]. This method can be applied to any set of quantitative variables, including MR measures [19]. Starting from these variables, additional features for each participant can be extracted from the graphs, namely measures of node centrality and node segregation, furnishing further information about the similarities between subjects which are expected to be higher among patients with the same disorder. The resulting graph-based features can be combined with the other features, to improve the classification accuracy for both binary and multi-class problems.

The aim of the present study is to discriminate patients with different forms of degenerative parkinsonisms using a multi-class classification SVM approach applied to brain MRI quantitative markers, selected algorithmically from among the most discriminative subset of features derived from morphometric, volumetric, 1H-MRS, DTI, and also graph-based derived from these, extending our previous work [17]. The novelty of the work lies in the combined use of graph-based features, feature selection, and “one-vs-all” “one-vs-one” binary classifiers, to implement the multi-class classification. The proposed approach is described in detail in the Materials and Methods.

2. Materials and methods

2.1. Subjects

Forty-seven PD patients (20 possible and 27 probable), 22 probable PSP (PSP-Richardson’s Syndrome variant [20]), 9 probable MSA-C, and 7 probable MSA-P consecutive patients who underwent brain MR at the Functional MR Unit of the Policlinico S. Orsola-Malpighi, Bologna, Italy, as part of their diagnostic work-up, were included in the study. All participants gave written informed consent to personal data processing for research purposes, and the study protocol was approved by the local Ethical Committee. Clinical diagnosis was performed by neurologists expert in movement disorders (PC, GCB, GG), according to current criteria [21–23]. Age at evaluation and sex were considered in the SVM model (2 features). We tested differences in demographic (age at evaluation, sex) and clinical (disease duration, Hoehn & Yahr modified scale scores) features using ANOVA followed by Bonferroni post-hoc analysis for continuous variables and χ² test for categorical variables (Table 1). Statistical significance was assumed for p value < 0.05. Analyses were performed with the program IBM® SPSS® v.21.

2.2. Brain MRI acquisition and analysis protocol

All participants underwent a standardized brain MR protocol with a 1.5 T GE Signa HDx 15 scanner. A quadrature bird cage head coil was used for signal reception. The protocol included 3D high-resolution T1-weighted fast spoiled gradient echo (FSPGR; TR = 12 ms, TE = 5 ms, 1 mm isotropic resolution), a coronal T2-weighted fluid-attenuated inversion recovery image (FLAIR; TE = 84.8 ms, TR = 8000 ms, 0.9375 mm resolution in plane, 3 mm slice thickness) and DTI (TR = 10000 ms, TE = 87.5 ms, 25 directions, b-value = 900 mm²s⁻¹, axial oblique FOV 32 cm, 1.25 mm reconstructed resolution in plane, 4 mm slices thickness) sequences. A single-voxel left cerebellar hemisphere 1H-MRS was acquired using the PRESS (Point REStolved Spectroscopy) sequence.

Data were analyzed and converted to numerical values as candidate features for use in the SVM-based classifiers described below. The selection of parameters was based on recent literature about the use of MRI and MRS for differential diagnosis of the parkinsonisms. We calculated median DTI parameters (Fractional Anisotropy, FA, and Mean Diffusivity, MD) in 42 Regions of Interest (ROIs) (84 features) and using a histogram-analysis method (34 features) [10]. Manual morphometric analysis was performed to measure midbrain area, pons area, the diameters of the Middle and Superior Cerebellar Peduncles (MCP and SCP), their ratios and the Magnetic Resonance Parkinsonism Index (MRPI) (7 features) [3,6,10]. Automated volumetric analysis was performed using FSL-FIRST tool (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/) on 3D T1-w images to calculate volumes of deep brain structures, lateral ventricles, cortical lobes, and cerebellum (21 features) [10]. (Additional details of MR data processing are provided in the Supplementary Materials.) Quantification of 1H-MRS-derived metabolites was performed by the fitting-program LCModel v 6.3. Ratios of N-Acetyl Aspartate (NAA), Choline (Cho), and myo-inositol (mI) to Creatine (Cr) and NAA/ml were calculated (4 features) [8].

2.3. Dataset modeling using graphs

A data vector consisting of all MR features from each patient was constructed and the Euclidean distance between each vector was used to construct a graph, a square matrix containing edges (similarity scores based on Euclidean distances) relating nodes (patients). Then, the minimum number (K) of desired edges for each node was fixed and two nodes were connected by an edge if either was among the K nearest neighbors of the other. Once the graph was built, six topological measures were calculated for each graph node and considered as additional features, namely node degree (the number of nodes to which a given node is connected), local average degree (which gives information about the degrees of the neighboring nodes of a given node), betweenness centrality and closeness centrality (which both measure the influence a node has on the other nodes), the eccentricity of a node (which is the greatest distance from a node to any other node) as a measure of segregation, and finally the clustering coefficient (which quantifies the number of connections that exist between the nearest neighbors of a node as a proportion of the maximum number of possible connections) [24].

2.4. Pattern recognition analysis

The main step of the analysis (Step 2 below) was performed by means of SVMs with probabilistic output [25], using the LIBSVM
library developed by Chang and Lin [26], and implemented using MATLAB. Other steps were performed by MATLAB scripts developed specifically for the study.

Step 0: data preparation. Six graph-based features were calculated as described above. Each quantitative variable was scaled to the interval [-1; 1] to give all features the same importance a priori.

Step 1: feature selection. For each classification problem considered steps 2-3 were repeated by constructing SVMs first on the entire set of features and then on a subset of features extracted by the ranking procedure described in Subsection 2.5.

Step 2: SVMs were constructed using only MR markers and excluding graph-based features, for multi-class classification, for which all the parkinsonisms were evaluated simultaneously, as either three classes (PD vs PSP vs MSA) or four, dividing MSA into MSA-C and MSA-P forms. In particular, to implement the multiclass classifier, a “one-vs-all” approach was applied. In this approach, several binary classifiers are trained, one tailored to each class. Each binary classifier produces a real-valued confidence score for its decision (a binary SVM with a probabilistic output was applied in the present work, following [26]). To classify an unseen sample, all the binary classifiers are applied to it, then the class predicted by the combined classifier is the one for which the corresponding binary classifier reports the largest confidence score. Compared to the “one-vs-one” approach (which trains a binary classifier for each pair of classes, then applies a voting scheme to predict the label of an unseen sample [27]), the “one-vs-all” approach has the advantage of using a smaller number of binary classifiers, and a larger number of training samples for each binary classifier. For completeness, and to allow comparison with previous studies, we also constructed binary “one-vs-one” classifiers with standard output. The results are presented in the Supplementary Materials.

Step 3: Step 2 was repeated using feature vectors that included graph-based features.

Step 4: statistical analysis. For each classification problem, a Leave-One-Out Cross-Validation (LOOCV) procedure was used to compute sensitivity and specificity of the solution, and thus the accuracy, computed as the ratio of the number of patients correctly classified to the total number of patients [28]. More precisely, one patient was left out from the training phase and the classifier was subsequently tested on this patient. This procedure was repeated for all the patients and the sensitivity, specificity, and accuracy were calculated averaging over the total number of patients. In addition, two simpler classifiers were created for each classification problem, in order to place the performance of the SVM in context. Firstly, we created single-feature linear SVMs for every non-graph feature, for each one-vs-all problem, and for each problem, we identified the best performing feature (or features) in terms of accuracy, along with its sensitivity and specificity. We also calculated a 0-feature classifier for every binary one-vs-all classification problem, that chose “all” for every case, to identify the irreducible minimum accuracy expected.

2.5. Feature selection

A feature selection method was applied to all the classification problems considered, to determine the best discriminant subsets of MR markers and graph-based features for diagnosing out-of-sample cases, and avoiding over-fitting. This was done in two steps. First the most informative individual features were identified, and then subsets of different sizes were tested for generalizability.

Each feature was ranked independently, using the relative entropy criterion [29], according to its ability to assign each patient to the correct disorder category, following the procedure described by Theodoridis and Kourtzoumbs [30], to then select subsets of 10, 20, 30 or 40 features used to create the SVMs described above. A priori, the best number of features to choose was unknown, so we selected the best performing subset for a given number of features, concentrating on relatively small numbers of features, given that one of the main aims of selection is to remove redundant features where possible. We decided not to consider datasets with a number of features more than 40 because we considered this number to be a good compromise between keeping the dimension of the problem relatively small and obtaining a good accuracy. The diagnostic accuracy of each set of selected features was evaluated using the Area Under the Curve (AUC) of the relevant Receiver Operating Characteristic (ROC) curve built by changing the threshold on the probabilistic output of the SVM. We estimated the diagnostic accuracy of the set of selected features by comparing the ROC curves obtained by using the entire set of features and by considering each subset. Reference [17] provides further details of all analysis methods employed.

We tested for significant differences in diagnostic properties between a classifier based on quantitative MR markers alone and after adding graph-based features to the dataset by comparing ROC curves with the method of DeLong et al. [31] using MedCalc v 16.1. This analysis allows for the comparison of features required for the selection process regardless of the specific threshold value of the SVM probabilistic output. For the evaluation of the results, on the other hand, a standard LOOCV was used, in order to compare the specific classifications obtained.

3. Results

In total data from 85 patients were included in the study. The characteristics of the subgroups in the sample differed, reflecting differences in the populations from which they were drawn (Table 1). In particular, the PSP patients were on average older than the other groups, and had a higher clinical severity score (both p < 0.01). Disease duration was greater in the MSA-C group compared to PD and PSP patient groups (p < 0.05).

Overall, SVMs solved multi-class problems discriminating each parkinsonism from all others with moderate-to-high accuracy (Fig. 1).
The MR features that best discriminated parkinsonianisms in the multi-class scenario were: MRPI, MD of the right MCP, MD of the right and then left putamen, MD of the right head of caudate nucleus, MD of the right then left pre-frontal WM, MD of the Posterior Fossa (PF) (25th percentile then median), median MD of the brainstem, median MD of the right, then left, then combined cerebellar hemispheres, and cerebellar volume.

4. Discussion

In the present study, we have demonstrated that pattern recognition analysis based on quantitative brain MR markers, obtained from both manual and semi-automated structural, microstructural and biochemical analyses can assist the diagnosis of parkinsonian disorders.

To reflect clinical practice as much as possible, the study examined all consecutive patients with the most frequent neurodegenerative parkinsonism in differential diagnosis with PD, i.e., PSP, MSA-C and -P, including patients at different disease stages, with various disease durations and ages at onset, and analyzed quantitative markers that can be obtained from clinical scanners. A binary SVM was applied in a “one-vs-all” setting, and multi-class SVMs with 3 or 4 classes were also tested, considering each disease simultaneously.

Feature reduction is an important part of a well-designed machine learning application. In the current study, feature selection was done both to identify the most informative features for classification, and to increase the generalization capacity of the resulting classifier, since it employs fewer features than a classifier based on all the features. This is especially important the case, typical in clinical applications, where the examples available to train the classifier are not numerous. Finally, another motivation for using feature selection is computational: a classifier based on fewer features classifies a new example more quickly. In order to evaluate the diagnostic usefulness of the proposed method at the single patient level, we applied a LOOCV procedure to test the generalization capability of the learning machine, increasing the validity of the approach. We found that, as expected, the best accuracy was achieved in most cases when only a subset of features was provided to the classifier. The results reported in Fig. 2 show that, in general, 10 or 20 features were enough to achieve results either as good as, or even better than, the ones obtained with the entire set of features. Overall, when graph-based features were not taken into consideration, the MR markers that were most frequently automatically extracted by the ranking procedure reflected alterations in brain regions that are usually evaluated to discriminate parkinsonisms in routine clinical practice, including infratentorial structures, basal ganglia and brain hemispheric white matter, and reflecting current understanding of the diagnostic accuracy of individual brain MR markers in differentiating parkinsonisms [10]. In the absence of the graph-based features, we found a very high specificity in discriminating PSP vs all (100%), which could be explained by the presence of diffuse alterations in the whole brain of these patients, and the relative low specificity of PD classification (84%), in line with the relative milder microstructural brain alterations in PD [10]. In all problems but one, a multi-feature improved upon the accuracy of single feature classification (Supplementary Table 1). For the MSA-P vs all problem, the MD of the middle
**Fig. 2.** Best discriminating features in the binary classification problems “one disease vs all” with and without the addition of graph-based features. The best discriminating features are color-coded as follows: black: not selected; white: selected; gray: not applicable; for DTI parameters: red: MD only; green: FA only; yellow: both MD and FA. DTI parameters are median values, except where stated as 25% pc. A complete list of features is shown in Supplementary Table 2.

MRPI: Magnetic Resonance Parkinsonism Index; MCP: middle cerebellar peduncle; SCP: superior cerebellar peduncle; PLIC: posterior limb of internal capsule; WM: white matter; pc: percentile; NAA: N-acetyl-aspartate; Cr: creatine; mI: myo-inositol.
cerebellar peduncles was able to classify cases with an accuracy of 0.94 compared to 0.92 for the best SVM. As classification without reference to the data (0-feature results) also gave an accuracy of 0.92, this suggests that the SVM was over-fitting the data. Note that the performance of the single-feature classifier was not itself satisfactory, with a sensitivity of 0.29. The addition of graph features to the SVM, by comparison, also yielded an accuracy of 0.94, but with a sensitivity of 0.6.

In the multi-class case, both considering the MSA separately and together, the accuracy of the classification was around 90% and the best discriminating features were once again structural alterations in the infratentorial structures, basal ganglia and hemispheric white matter, confirming previous literature [7,10].

It is worth mentioning that a few other recent studies have applied a pattern recognition approach to structural and functional brain MRI data in parkinsonian disorders. Several recent studies have applied SVM methods to the differential diagnosis of parkinsonian using MRI [13–16], but without the use of MRS techniques. Two studies are most directly comparable to the current study.

Using both binary and multi-class SVM [15] 3D T1 data derived from Voxel Based Morphometry (VBM) and atlas-based volumetry respectively to differentiate PD, PSP and MSA, accuracies were found to be >80% except for PD vs HC and MSA-C vs MSA-P (<70%).

Marquand et al. [16] applied four different multi-class SVM-based classifiers to 3D T1-weighted images obtained from a group of 60 parkinsonisms and 19 HC. They performed analysis both on the whole brain and in target regions of interest, i.e., the “subcortical motor network” (cerebellum, midbrain, SCPs decussation, caudate, putamen, pallidum and accumbens nuclei), finding that these structures discriminated among parkinsonisms better than the whole brain analysis with moderate to high accuracy. Voxel-based discriminated PSP were localized in midbrain, decussation of the SCPs and caudate nuclei, and in the cerebellar cortex, MCPs and brainstem in the case of MSA. Consistent with our findings, they found that the performance of the classifier was higher when considering MSA-P and MSA-C together, with an accuracy of 91.7% for a three-class classifier [16]. This compares with the 94% accuracy score that we reached using graph-derived features. We should note that we included more modalities in our original feature set including, for the first time, a 1H-MRS-derived biochemical cerebellar profile. Note that the results of the binary ‘one-vs-one’ classifier comparison are discussed separately in the Supplementary Materials.

Compared to previous SVM-based studies of parkinsonism, our study also evaluated the contribution of graph-based features extracted from the original brain MR quantitative markers. This approach has proved to be useful for the identification of connections and similarities between patients with the same disorder, and it was applied for early detection of Alzheimer’s disease [19], but has not been tested on the parkinsonism to date. In fact, modeling the MR markers by means of a graph can be useful to extract features that better highlight the connection, and thus the similarity, between patients with the same form of parkinsonism. In the current study, results obtained when including graph-based features in the SVM improved accuracy in three out of five binary classification problems, compared to when such features were not used, and reduced the number of features needed to achieve that accuracy, as previously reported for other applications [19].

In particular, we could discriminate PD vs other parkinsonisms with high sensitivity and specificity (94% and 95%, respectively). For the multi-class problem, the inclusion of graph-derived features improved the accuracy for the 3-class classification problem, which considered MSA as a single class.

Looking beyond the binary classification problems, we have also demonstrated the feasibility of using SVMs to create a tool for the differential diagnosis of all the parkinsonian disorders, which should be far more useful in practice, as it could be integrated directly into the clinical decision-making process. We expect that the form of classifier proposed here, that compares the outputs of different “one-vs-all” binary classifiers – based on efficient automatically selected features, including graph-based features - to build the multi-class classifier, will prove useful in assisting clinical diagnosis, although some limitations must be acknowledged.

The main limitations of the study are represented by the limited size of the sample, in particular of MSA patients. The relatively modest advantage presented by the inclusion of graph-based features is likely due to the limited sample size, since this technique should present advantages in databases with at least hundreds of patient for each category. The lack of post-mortem definite diagnosis in all cases should also be acknowledged. Moreover, further prospective studies on patients with unclassifiable parkinsonism at onset will be necessary in order to support the diagnostic usefulness of this approach in the early stages of the diseases.

In conclusion, the results obtained in the paper show that SVMs applied to multimodal brain MR imaging data sets and using graph-based features represent a new and accurate approach to discriminate parkinsonisms, and one that, once training samples reach a sufficient size to allow a robust generalizability of reported accuracies, should prove to be a useful tool to assist diagnostic work-up.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.parkreldis.2017.11.543.

References

from Parkinson’s disease, Mov. Disord. 23 (16) (2008 Dec 15) 2370–2376.


D.J. Cook, I.B. Holder (Eds.), Mining Graph Data, Wiley, 2006.


