Complex networks reveal early MRI markers of Parkinson's disease

Nicola Amoroso[a,b], Marianna La Rocca[a,b,*], Alfonso Monaco[b], Sabina Tangaro[b], Roberto Bellotti[a,b]

[a] *Dipartimento Interateneo di Fisica "M. Merlin", Università degli studi di Bari "A. Moro", Italy*
[b] *Istituto Nazionale di Fisica Nucleare, Sezione di Bari, Italy*

Abstract

Parkinson's disease (PD) is the most common neurological disorder, after Alzheimer's disease, and is characterized by a long prodromal stage lasting up to 20 years. As age is a prominent factor risk for the disease, next years will see a continuous increment of PD patients, making urgent the development of efficient strategies for early diagnosis and treatments. We propose here a novel approach based on complex networks for accurate early diagnoses using magnetic resonance imaging (MRI) data; our approach also allows us to investigate which are the brain regions mostly affected by the disease. First of all, we define a network model of brain regions and associate to each region proper connectivity measures. Thus, each brain is represented through a feature vector encoding the local relationships brain regions interweave. Then, Random Forests are used for feature selection and learning a compact representation. Finally, we use a Support Vector Machine to combine complex network features with clinical scores typical of PD prodromal phase and provide a diagnostic index. We evaluated the classification performance on the Parkinson's Progression Markers Initiative (PPMI) database, including a mixed cohort of 169 normal controls (NC) and 374 PD patients. Our model compares favorably with existing state-of-the-art MRI approaches. Besides, as a difference with previous approaches, our methodology ranks the brain regions according to disease effects without any *a priori* assumption.

Keywords: Parkinson's disease, MRI, Complex Networks, Machine learning

*marianna.larocca@ba.infn.it (Marianna La Rocca)
Email address: nicola.amoroso@ba.infn.it (Nicola Amoroso)

1. Introduction

Parkinson's disease (PD) is a heterogeneous progressive neurological disorder, firstly described almost two centuries ago, basically related with early death of dopaminergic neurons in the substantia nigra and characterized by both motor and non-motor features (Gibb and Lees, 1988; Jankovic, 2008). It is recognized that age is the greatest risk factor for PD, its incidence reaches a maximum at about 80 years of age, thus the rising life expectancy is expected to increase the number of patients at more than 30% by 2030 (Dorsey et al., 2007).

The slow progression is one of the most important features of PD. The disease course can be roughly separated in two phases; the first prodromal, usually named "premotor", phase can last up to 20 years and accounts for symptoms such as: impaired olfaction, constipation, depression, rapid eye movement sleep behavior disorder (RBD) and excessive daytime sleepiness (EDS) (Singaram et al., 1995; Gagnon et al., 2002; Chaudhuri et al., 2006). In fact, these symptoms could double the individual's risk of developing the disease.

The second phase is early characterized by insurgence of typical bradykinesia, tremor and fatigue; then in the advanced stages by psychosis, dysphagia, freezing of gait, falls and postural instability (Friedman and Friedman, 1993; Huber et al., 1986). It is known that the average latency between the onset of prodromal and motor symptoms is about $12 - 14$ years (Postuma et al., 2012). Thus, it would be of paramount importance the development of diagnostic strategies able to detect the disease in its prodromal phase and outline efficient markers. As shown in (Kalia and Lang, 2015), a prominent role in early diagnosis should be played by RBD, EDS, Hyposmia, depression and mild cognitive impairment (MCI) which are typical of the non-motor, prodromal or very early, PD phase, see Figure 1.

For what concerns imaging, candidate markers include positron emission tomography (PET) (Antonini et al., 1997; Hansen et al., 2016; Masdeu, 2017) or single photon emission computed tomography (SPECT) (Hirschauer et al., 2015; Suwijn et al., 2015; Adeli et al., 2017) methods. In fact, these methodologies can accurately detect PD. However, these methodologies are based on the detection of substantial losses of dopaminergic neurons, for example in the substantia nigra, whilst it would be desirable to diagnose the disease

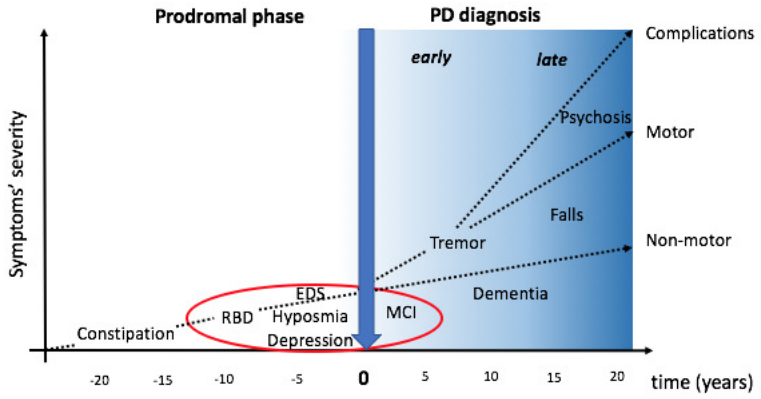

Figure 1: PD diagnosis is related to the onset of motor symptoms (time 0). The symptoms characterizing the prodromal phase and the years immediately following the diagnosis are: Rapid eye movement sleep Behavior Disorder (RBD), Excessive Daytime Sleepiness (EDS), Hyposmia (a reduced ability to smell and to detect odors) and mild cognitive impairment (MCI). Accordingly, these symptoms are usually enclosed in models trying to forecast the disease onset.

before this degeneration has occurred in order to enable early diagnosis before the onset of motor symptoms. With this regard, it should be taken into account that PD patients could lose up to 80% of dopamine before symptoms appear (Miller and O'Callaghan, 2015).

With this regard, the definition of new markers will play a fundamental role. It is clear that a single marker will not be able to allow accurate diagnosis and monitor disease progression. Instead, a combination of different markers should provide a more reasonable approach. As previously mentioned, PET and SPECT measures are very effective in the motor phase, and they may support the diagnosis as well as monitor disease severity and progression. Transcranial sonography is another promising approach whose clinical applicability is still controversial (Bouwmans et al., 2013; Pilotto et al., 2015).

MRI markers could monitor structural changes in the brain and suggest increased risk for PD (Salvatore et al., 2014) or be employed for differential diagnosis of PD syndromes (Duchesne et al., 2009; Marquand et al., 2013; Chagas et al., 2017). For example, voxel based morphometry (VBM) has revealed significant gray matter reductions in PD patients with dementia (Summerfield et al., 2005); MRI has also shown progressive atrophy in PD, an effect already detectable in the early stage of the disease (Beyer et al.,

2007; Tessa et al., 2014). Thus, it would seem that there is still room to define effective MRI markers which outline the disease process before the death of dopaminergic neurons has triggered irreversible damages.

Several works have investigated the applicability of machine learning strategies to MRI data with fluctuating outcomes. (Focke et al., 2011) tried without success to use VBM features for individual classification using a Support Vector Machine (SVM). However, (Cherubini et al., 2014) demonstrated that VBM features combined with diffusion tensor imaging can effectively distinguish PD patients from subjects with progressive supranuclear palsy. More recently, a synergistic paradigm combining Kohonen self organizing map and SVM claimed that MRI features can reach accurate classification performances including subjects with no dopaminergic deficit (Singh and Samavedham, 2015). Feature selection strategies seem to play a relevant role to define imaging markers accurately distinguishing PD patients from controls (Adeli et al., 2016). These different approaches share a not negligible feature, all of them rely on the supervised selection of PD-related regions of interest to obtain statistically significant associations between anatomy and clinical phenotype. Although these approaches reach excellent results, they can be limiting as they prevent the investigation of novel brain regions.

In this paper, we use MRI data from the Parkinson's Progression Markers Initiative (PPMI) to extract imaging markers and learn an accurate classification model. With this goal, we introduce a brain connectivity model basing on gray matter and white matter voxel distribution. The proposed approach adopts a brain patch segmentation, thus it avoids common drawbacks of voxel-wise approaches, e.g. the lack of significance due to high dimensionality of the feature space. Besides, this methodology does not depend on fully-automated brain segmentation algorithms, whose accuracy could be poor, for region of interest definition. We measure how different brain regions are correlated and for each region we measure simple topological quantities; accordingly, we build a model including atrophy effects locally induced by the disease and accounting for whole-brain modifications thanks to the network framework.

Complex network approaches have already proven their effectiveness in several cases, also in neuroimaging applications. MRI provides a useful base of knowledge when considering the topological organization of the brain. In fact, findings from structural (without excluding the functional) graphs point to a loss of highly connected areas in several brain diseases (Tijms et al., 2013). Graph theory provides two important methodological insights.

Firstly, it associates to each node quantitative measurements characterizing its role and importance within the network; secondly, it enables a direct description of the whole brain from a global perspective, thus letting emerge properties which affect the brain as a system (Bullmore and Sporns, 2009).

Several examples can be found for Alzheimer's disease. (Stam et al., 2007) investigated small-worldness properties and found that diseased brains show a loss of connectivity. Also, the topological organization of the brain itself could be used as a marker; in fact, an increment of the shortest path length could denote an impaired organization (Lo et al., 2010). Another approach to detect the impairment of connectivity consists in measuring node-related quantities, as for example the rich-club property (Daianu et al., 2014). Finally, in previous works (La Rocca et al., 2017; Amoroso et al., 2017), we showed how complex network measures can be used to characterize Alzheimer's disease. Of course, PD has its specificities, for example it cannot be considered a disconnectivity disease. However, as previously mentioned structural changes could be useful markers. It would be interesting to evaluate whether PD patients show significant brain structural changes, both locally and globally, and whether complex network measures can capture these effects. In this work, basing on our previously mentioned works, we introduced a novel machine learning approach to combine network and clinical features within a unique classification score. Besides, we demonstrated its effectiveness on PD; as a matter of fact, we could not find any PD study investigating the adoption of complex network measures.

This work offers three main contributions: (i) we propose an unsupervised general methodology to model brain connectivity for both healthy subjects and patients; (ii) we explore which regions are significantly affected by the disease; (iii) we propose a novel learning strategy to combine network and clinical features; (iv) we define an accurate diagnostic tool for PD diagnosis basing only on MRI features; (v) we highlight the existence of an optimal scale to study PD. It is worth mentioning that previous approaches exploited *a priori* definition of regions of interest within the brain and therefore they could suffer a loss of information. In brief, the proposed approach can learn an accurate model to discriminate controls and patients and, eventually, detects possible novel imaging markers of the disease. Therefore, the use of MRI features becomes strategic for the development of early diagnosis tools or a better characterization of PD in its early stages.

2. Materials

Data used in preparation of this work was obtained from the Parkinson's progression markers initiative (PPMI) database[1] (Marek et al., 2011). MRI data acquired by the PPMI for this study consisted of MPRAGE T1 brain scans from 3T SIEMENS MAGNETOM Trio scanners. Images were acquired with the following parameters: repetition time 2300 ms, echo time 2.98 ms, flip angle 9° and voxel size $1 \times 1 \times 1$ mm^3, so that the equivalence 1 voxel = 1 mm^3 holds.

The PPMI is a clinical study whose main goal is the identification of PD markers in order to enhance the comprehension of the disease and eventually help the development of disease modifying therapies. The PPMI includes a mixed cohort of normal controls (NC) and PD patients; the database also includes subjects without dopaminergic deficits, namely SWEDD, that are disregarded in this work. Data from PPMI comes from different worldwide sites, along with structural MRI it is possible to find other imaging modalities such as SPECT or demographic and clinical metadata, such as age, gender and cognitive scores. NC subjects are both age- and gender-matched with the PD patients. It is worth noting that PD patients are *de novo* patients in that they are unmedicated, an important aspect as PD therapies could not have the desired effect of modifying the possible markers. More importantly, PD patients enrolled are mostly at the first stages of the disease, according to the Hoehn and Yahr scale (Hoehn et al., 1998); in fact, (Marek et al., 2011) explains that 98% of the subjects affected by the disease is in stages 1 and 2 (over 5), corresponding to mild inconvenient without disabling symptoms.

The database consisted of two populations including respectively 107 male and 62 female NC, for a total of 169 subjects, and 243 male and 131 female patients, for an overall amount of 374 PD subjects. The populations are matched for age (60.2 ± 11.5 for NC and 61.6 ± 9.8 for PD). In the following Table 1 the baseline values for the Epworth Sleepiness Scale (ESS), the Geriatric Depression Scale (GDS), the Montreal Cognitive Assessment (MoCA), the Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) and the Rapid eye movement sleep Behavior Disorder (RBD) are enlisted.

ESS is a standardized simple measure for sleep propensity (Johns et al., 1991). Daytime sleepiness is usually associated to sleep disorders, but it can

[1] http://www.ppmi-info.org/data.

Diagnosis	Age	ESS	GDS	MDS-UPDRS	MoCA	RBD
PD (374)	61.6 ± 9.8	3 [2, 6]	2 [1, 3]	29 [18, 41]	27 [26, 29]	5 [3, 8]
NC (169)	60.2 ± 11.5	2 [1, 4]**	1 [0, 2]**	27 [17, 37]*	28 [27, 29]**	5 [3, 8]

Table 1: Demographic and clinical information. Mean age and standard deviation are given, for other indicators with asymmetric distributions medians and interquartile ranges are preferred. Significant differences between normal controls (NC) and Parkinson's disease (PD) are reported with the Kruskal-Wallis p-value (* for $p < 0.05$ and ** for $p < 0.01$).

also be a symptom of prodromal PD phase along with RBD. Like constipation and olfactory disturbance, RBD can precede the development of the motor signs of Parkinson's disease and longitudinal data suggest that RBD heralds the onset of motor symptoms in up to 40% of patients (Chaudhuri et al., 2006). RBD is measured according to the screening questionnaire proposed in (Stiasny-Kolster et al., 2007). ESS and RBD scores lie within ranges of normalcy for both NC and PD subjects, however PD patients showed a small but significant increment in ESS (Kruskal-Wallis p-value < 0.01).

GDS score is based on a questionnaire of 30 items with binary outputs (Yesavage and Sheikh, 1986). For each affirmative answer 1 point is scored; healthy people should score 5 ± 4, mildly depressed and very depressed people 15 ± 6 and 23 ± 5 respectively. In particular, for the present study, the shorter form was used (Yesavage et al., 2000). Accordingly, the subjects of the study were neither mildly nor very depressed, nevertheless PD patients showed a significantly higher GDS score (Kruskal-Wallis p-value < 0.01).

The motor symptoms are taken into account by the Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) (Goetz et al., 2008). It is important to recall how the MDS-PD criteria assign a central role to motor symptoms to define clinical PD (Postuma et al., 2015); however, also non-motor manifestations are present in many patients so that the related indicators can play a fundamental role for diagnosis, even though they cannot capture the complexity of this heterogeneous disease. As expected, MDS-UPDRS significantly distinguishes (Kruskal-Wallis p-value < 0.05) the NC and PD cohorts.

Mild cognitive impairment (MCI) is a symptom commonly found in PD patients; it usually occurs with the onset of motor symptoms and may be a harbinger of dementia. MCI condition could be related to early PD symptoms, such as RBD, in any case it is known that it may be found in up to 80% of long term PD patients (Litvan et al., 2012). The Montreal Cognitive Assessment (MoCA) index is the preferred measure for accurate screening of

cognition in PD (Dalrymple-Alford et al., 2010). A final MoCA score of 26 and above is considered normal: PD and NC cohorts of this study resulted normal on average with 27.0 ± 2.3 and 28 ± 1 respectively. A small but significant difference was observed (Kruskal-Wallis p-value < 0.01).

3. Methods

The proposed approach aims at using MRI data to extract novel and efficient PD markers for early diagnosis. MRI scans from PPMI were processed to be both intensity and spatially normalized. Then, we introduced a connectivity model for each brain and obtained a feature representation using measures derived from the network description. Finally, we used the feature representation to learn a supervised classification model within a nested cross-validation framework. Besides, as the NC and PD classes are not balanced, we performed a stratified cross-validation by granting for each round that the same number of subjects was sampled for the two classes. The classification consists of three distinct steps: firstly, Random Forest classifiers (Breiman, 2001) are used as a wrapper for feature selection; then, the important features are used to feed, within the same cross-validation cycle, a second RF classifier in order to obtain a classification score for each subject; finally, an SVM classifier (Cortes and Vapnik, 1995) combines these scores and the other clinical features to discriminate the NC and PD classes. A schematic overview of the method is shown in the following Figure 2.

The main goal of the method is to provide a classification score for PD. Besides, the methodology can be used to investigate which regions are mostly affected by the disease and rank them according to statistical significance.

3.1. Image processing and network construction

MRI scans were skull stripped and aligned with an affine registration using the FSL library developed by the Oxford Centre for Functional MRI of the Brain (FMRIB), specifically the Brain Extraction Tool (BET) (Smith, 2002) and the FMRIB's Linear Image Registration Tool (FLIRT) (Jenkinson and Smith, 2001; Jenkinson et al., 2002) were employed. Both the procedures were performed with default parameters, the reference image used for registration was the T1 MNI152 template with $1 \times 1 \times 1$ mm^3 voxel size. Then, the scans were segmented in rectangular non overlapping boxes (from now onwards *patches*) of volume $V = l_1 \times l_2 \times l_3$ using the medial longitudinal fissure to separate the left and right hemispheres. Accordingly, each

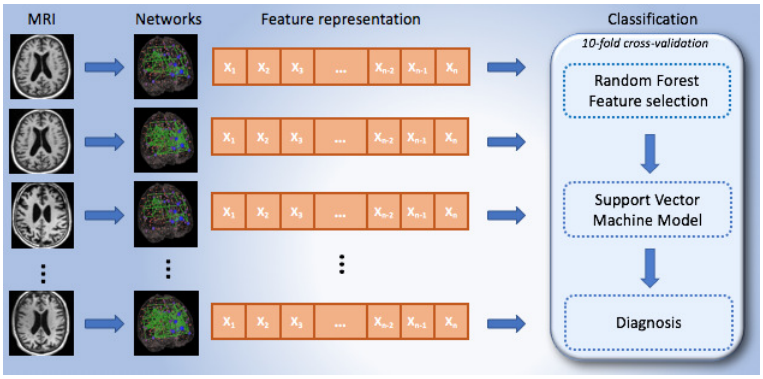

Figure 2: Methodology flowchart. MRI scans are processed to obtain a network representation, in particular they are preliminarily registered to the MNI152 brain template with an affine transformation. Then, for each node several features are computed and a feature representation is obtained. These features evaluate the node importance within the networks. The most informative features are selected with Random Forest wrapper and summarized in a classification score, then a Support Vector Machine combines these score with clinical features to distinguish NC and PD groups. The entire classification process is performed in 10-fold cross-validation.

hemisphere was divided in an equal number of patches. These patches were considered the nodes of a weighted network having as weights the pairwise Pearson's correlation measured between each pair of nodes throughout the whole brain.

The number of patches, a brain can be divided into, depends on the volume V. As typical normalized volumes of substantia nigra range from 153 to 221 mm^3 (Kwon et al., 2012), we chose V = 125 mm^3 with $l_1 = l_2 = l_3 = 5$ mm. Thus, the resulting networks consisted of 12219 nodes, each node of the network consisting of a patch including 125 voxels. To evaluate the presence of a link between two nodes we represented each patch through a 125−dimensional vector and measured the pairwise Pearson's correlation. This measure emphasizes how similar two patches are, indeed, two patches are highly correlated if, voxel by voxel, their gray levels are similar and therefore if, spatially the distributions of white matter, gray matter and cerebrospinal fluid (corresponding to voxel intensities of decreasing intensity) are similar as well.

Correlations were measured in absolute value, thus disregarding left-right symmetries, but keeping intact the informative content about structural modifications.

Thus, for each subject we introduced a complex network model. This model was investigated with measures borrowed by graph theory, particularly concerning weighted graphs. In order to remove noisy connections and avoid as far as possible the loss of strategical links, we thresholded the networks disregarding connections corresponding to less than moderate correlations ($|r| < 0.3$). A study on how the threshold affects the capability of the network features to reveal pathological changes is reported in **??**.

3.2. Feature representation

The underlying hypothesis of the proposed approach is that structural changes of the brain, measured by correlations, affect the connectivity patterns. We expect that these changes mostly concern (i) the intensity of the connections of a node, (ii) the number of connections a node has and (iii) which nodes it is connected with.

The first assumption holds because atrophic changes affecting a brain region tend to weaken the correlations of that specific region with other healthy GM/WM regions and enforce correlations with other atrophic regions or regions containing mostly CSF. As GM, WM and CSF are not evenly distributed, these changes should be detectable. The second assumption stands with the first one; as the intensity of connections changes, the number of connections must change too. Finally, the third assumption is a direct consequence of the first two: modifying the intensity of connections and the number of connections is equivalent to remove some links and create new ones. Accordingly, the organization itself of the networks should change; these effects can be detected with some specific complex network measures.

The intensity of connections of a network is, by definition, the strength of the network nodes. In order to detect how these structural changes of the brain are locally distributed, we considered the strength s_i of each node i of the network, *i. e.* the intensity of connections of a single patch.

Strength provides insight on the intensity of the connections of a particular brain region, however without taking into account if the number of connections is preserved. Of course, this second aspect should not be neglected, in fact it is in principle, possible, that a node i preserves its strength even losing or acquiring some connections, provided that the sum of the weights remains unchanged. Thus, to detect this effect, we measured, in addition to the strength, the inverse participation ratio Y_i^{-1} which evaluates how unevenly the weights of the links of the node i are distributed (Menichetti et al., 2014):

$$Y_i = \sum_{j=1}^{N} \left(\frac{w_{ij}}{s_i} \right)^2 . \tag{1}$$

to detect variations in brain connectivity among nodes with the same degree k, which is the number of connections existing upon a node, we evaluated also the conditional values of strength $s(k)$ and inverse participation ratio $Y^{-1}(k)$ for each subject and for each degree k ranging from 1 to 12219:

$$s(k) = \frac{1}{N_k} \sum_{i=1}^{N} s_i \delta(k_i, k); \tag{2}$$

$$Y(k) = \frac{1}{N_k} \sum_{i=1}^{N} Y_i \delta(k_i, k); \tag{3}$$

where N_k is node number having degree k and δ is the Kronecker function that is 1 when a node i has degree k and 0 otherwise. These measures relate the intensity of the connections and their importance in the degree distribution of the single subject networks. Finally, to better capture inter-subject variations we considered the degree distribution of the whole training cohort k^{global}, an array whose elements k_i^{global} indicate the number of links connected to a node i over all the training subjects and we obtained the previous 4 quantities s Y^{-1} $s(k)$ and $Y^{-1}(k)$ weighing them on the global degree k^{global} and thus according to an overall perspective:

$$s'_i = \sum_{j=1}^{N} w_{ij} k_i^{global}; \tag{4}$$

$$Y'_i = \sum_{j=1}^{N} \left(\frac{w_{ij}}{s_i} \right)^2 k_i^{global}; \tag{5}$$

$$s'(k) = \frac{1}{N_k} \sum_{i=1}^{N} s'_i \delta(k_i, k); \tag{6}$$

$$Y'(k) = \frac{1}{N_k} \sum_{i=1}^{N} Y'_i \delta(k_i, k). \tag{7}$$

Thus, we obtained a 8-dimensional feature representation: s, Y^{-1}, s' and Y'^{-1} for each node and $s(k)$, $Y^{-1}(k)$, $s'(k)$ and $Y'^{-1}(k)$ for each degree.

3.3. Feature selection and Classification

Removing null mean and variance features and highly correlated features with a threshold of 0.7 from the feature representation obtained, 4048 features remain. For the feature reduction, we chose a moderate correlation threshold in order to reduce the great mole of features without losing information. Using all the original features as input into Random Forest, besides being computationally intensive and time consuming, would not allow us to reduce high dimensionality of the model to increase its generalization and give a deep insight into underlying mechanisms. After the first feature reductions the still elevated dimensionality of the model required a higher future selection level. There are three distinct approaches one could choose to tackle this task (Saeys et al., 2007): filter methods are fast and scalable, but they ignore possible interactions among the features; wrapper methods are computationally intensive in that they explore the space of features by evaluating random subsets and using supervised classifiers to find optimal configurations, but they are able to take into account feature interactions; embedded techniques are those for which the search for best discriminating features is built within the model.

We chose a hybrid approach in that we used Random Forests for feature selection but not for the model. Random Forests are an ensemble of classification trees, whose trees are grown by bootstrapping training data and randomly selecting at each split a candidate set of features. Given f features, at each split $\sqrt{f}$ features are randomly picked and each tree is grown unpruned to obtain low-bias models; the main idea behind Random Forests is the use of random variable selection resulting in low correlation of single trees; as a consequence, the overall classifier yields a model that can achieve both low bias and low variance (Breiman, 1996).

Random Forests are particularly suitable for the present analysis where the number of variables exceeds the observations (Díaz-Uriarte and De Andres, 2006). Moreover, it is a robust and easy-to-tune model, it does not overfit thanks to internal bagging and, more importantly, it provides a continuous measure of feature importance. For our experiments we used the implementation provided by *R version* 3.2.2 with the package *randomForest version* $4.6-10$; a standard configuration was used: each forest was grown with 500 trees.

We adopted a hierarchical combination of two forests. For each 10-fold cross-validation round, the first forest, for each split, randomly picked $\sqrt{f} = \sqrt{4048}$ features and selected a number of features, which were on

average 60. Features found by the first classifier were chosen in order to exceed in importance the third quartile of importance distribution computed in terms of mean accuracy decrease. Then a second forest within the same cross-validation cycle, was trained with these selected features to obtain a classification score. The results of this procedure were twofold: on one hand we selected the most discriminative features and on the other hand we summarized the whole information content provided by the complex network measures in a unique score. Thus, this score outlined which subjects had shown brain structural and topological changes significantly associated to the diagnosis. The presence of brain regions with modified centrality, the disrupt of connectivity, along with the previously mentioned properties concerning for example the intensity or the uniformity of connections, were all included in the classification score.

We designed this approach as the number of features deriving from the complex network description can be overwhelming if compared to the available clinical features of Table 1. It is important to remind that for each subject clinical features are provided at the baseline. Accordingly, a diagnostic model relying on clinical features of prodromal or early PD phase and structural MRI data was designed. Finally, for the discrimination of PD patients and normal controls, we trained a third radial Support Vector Machine (SVM) classifier combining the classification score and the clinical features and keeping fixed the training and validation set of the previous 10-fold cross-validation. For our analysis we used the *R* package *e1071 version 1.6-7* with the default implementation (cost = 1 and gamma = 0.003).

4. Results

4.1. Classification performance

The proposed methodology both detects which regions are mostly affected by the disease and uses the network measures to provide a classification score. Besides, the use of clinical features concerning the PD prodromal phase or the disease onset can support the early diagnosis. In order to evaluate the effectiveness of the proposed procedure we used standard machine learning techniques, such as the previously mentioned Random Forests and SVM algorithms. All the presented results were acquired with a 10-fold cross-validation framework. We measured the informative content provided by our complex network approach combined with the clinical features training an SVM classifier. Besides, we separately evaluated the informative content of respectively

the network measures, using the classification score of a Random Forest (RF) classifier, and the clinical features, training an SVM classifier. The results are shown in Figure 3.

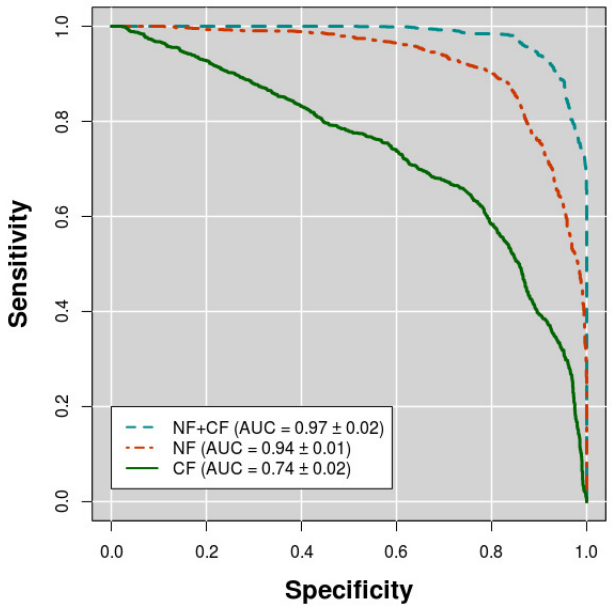

Figure 3: Classification performance in terms of area under the receiver operating characteristics curve (AUC). The combined use of network features (NF) and clinical features (CF) reaches the highest AUC $= 0.97 \pm 0.02$ (cyan dashed line). CF on their own reach an AUC $= 0.74 \pm 0.02$ (dark green continuous line), which is reasonable for baseline subjects whose symptoms are mild. NF provide effective markers for PD, in fact basing on these feature it is possible to diagnose PD with an AUC $= 0.94 \pm 0.01$ (dark red dash-dot line).

The combined use of MRI and clinical features gives the best performance as summarized in Table 2.

Area under the receiver operating characteristics $AUC = 0.97 \pm 0.02$, accuracy $ACC = 0.93 \pm 0.04$, sensitivity $sens = 0.92 \pm 0.06$ and specificity $spec = 0.93 \pm 0.07$. These results were significantly higher than those obtained by using only the complex network measures; in fact, we found in this case $AUC = 0.94 \pm 0.01$, $ACC = 0.88 \pm 0.06$, $sens = 0.85 \pm 0.09$ and $spec = 0.88 \pm 0.09$. These results were averaged on 1000 cross-validation rounds

features	AUC	ACC	sens	spec
Network measures	0.94 ± 0.01	0.88 ± 0.06	0.85 ± 0.09	0.88 ± 0.09
Clinical scores	0.77 ± 0.01	0.70 ± 0.08	0.65 ± 0.12	0.75 ± 0.11
Both	$\mathbf{0.97 \pm 0.02}$	$\mathbf{0.93 \pm 0.04}$	$\mathbf{0.93 \pm 0.06}$	$\mathbf{0.92 \pm 0.07}$

Table 2: NC vs PD classification performances for feature typology. Area under the receiver operating characteristics (AUC), accuracy (ACC), sensitivity (sens) and specificity (spec) are reported with the relative standard deviations. Best performance (bold) is obtained with a combined use of network and clinical features.

and significance was assessed with z-tests; for all comparisons we found 1% significance.

Clinical features resulted in a classification performance significantly lower than those previously reported. Thus, the information content provided by the proposed model gives a significant contribution. In particular, we found when using only the clinical features: $AUC = 0.77 \pm 0.01$, $ACC = 0.70 \pm 0.08$, $sens = 0.66 \pm 0.12$ and $spec = 0.73 \pm 0.11$. Figure 4 allows us to appreciate this effect from a different perspective.

Classification scores based only on clinical features consistently tend to overlap and assign to PD subjects low scores. In fact, sensitivity, which is basically the discriminative power for positive subjects, is lower than specificity. On the contrary, the discrimination of the two classes is greatly enhanced when introducing complex network markers.

4.2. Comparison with standard methods

In order to assess the effectiveness of the proposed approach, we compared its classification accuracy with that of two standard approaches. In particular, we used FreeSurfer Fischl (2012) to extract some structural features, such as grey matter and white matter volumes of subcortical brain structures or the average cortical thickness of specific regions for a total of 181 ROI features. Then, we performed a standard VBM pipeline Ashburner and Friston (2000) to detect voxels showing a significant (p-value < 0.01) association with the diagnosis, these voxels provided another feature representation. Both, the ROI and the VBM descriptions were used to feed the classification framework previously described. In this way, we obtained a direct comparison evaluating the informative power of the proposed methodology and ROI/VBM approaches, see Figure 5.

The proposed method (AUC $= 0.97 \pm 0.02$) compares favorably with VBM (AUC $= 0.93 \pm 0.04$) and ROI (AUC $= 0.82 \pm 0.06$) descriptions. It is worth

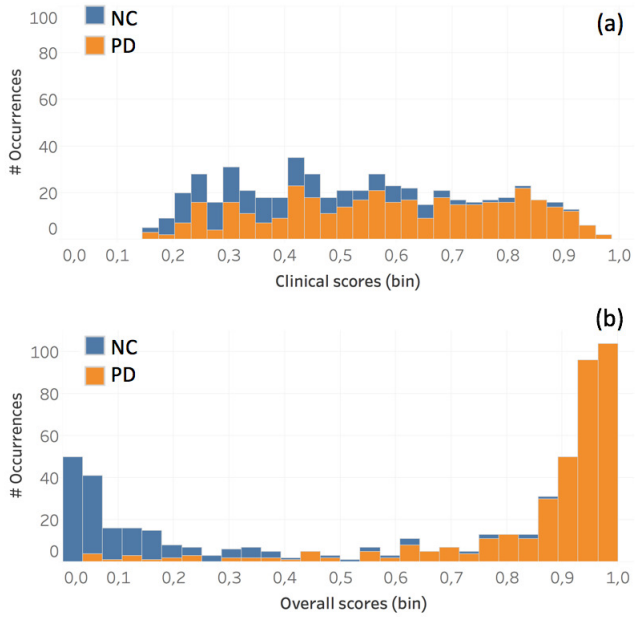

Figure 4: The classification score distribution using (a) only clinical features and (b) combining them with network measures. Each column of the histogram contains the number of NC subjects (blue) and PD patients (orange) whose score lies in that bin. Classification scores obtained using both network and clinical features show a greatly enhanced class separation.

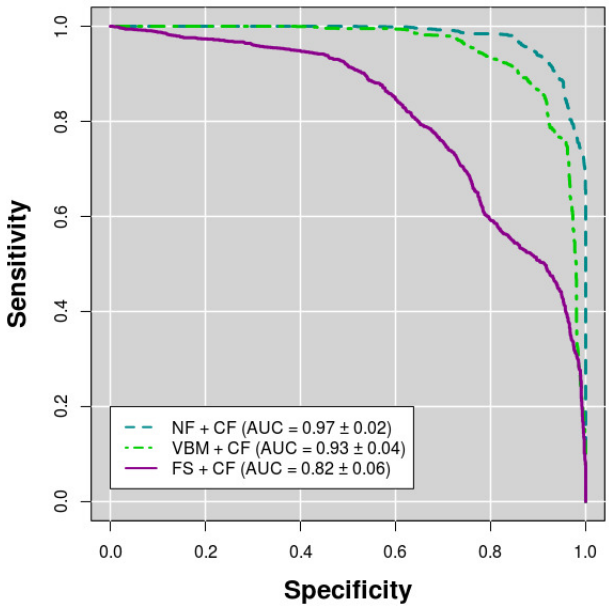

Figure 5: Classification performance in terms of area under the receiver operating characteristics curve (AUC). The combined use of network features (NF) and clinical features (CF) reaches the highest AUC $= 0.97 \pm 0.02$ (cyan dashed line). The combination of Voxel Based Morphometry (VBM) features and CF reaches an AUC $= 0.93 \pm 0.06$ (light green dash-dot line). Finally, FreeSurfer features (FS) combined to CF give an AUC $= 0.82 \pm 0.06$ (magenta continuous line).

mentioning that even when not considering clinical features, the network description remaines the most effective (AUC $= 0.94 \pm 0.01$) both in respect of VBM (AUC $= 0.87 \pm 0.05$) and ROI (AUC $= 0.70 \pm 0.06$) approach. A summary of this comparison is presented in Table 3.

features	**AUC**	**ACC**	**sens**	**spec**
NF + CF features	$\mathbf{0.97 \pm 0.02}$	$\mathbf{0.93 \pm 0.04}$	$\mathbf{0.93 \pm 0.06}$	$\mathbf{0.92 \pm 0.07}$
VBM + CF	0.93 ± 0.04	0.86 ± 0.06	0.88 ± 0.08	0.86 ± 0.08
FS + CF	0.82 ± 0.06	0.72 ± 0.07	0.74 ± 0.10	0.71 ± 0.12
NF	0.94 ± 0.01	0.88 ± 0.06	0.85 ± 0.09	0.88 ± 0.09
VBM	0.87 ± 0.05	0.79 ± 0.08	0.77 ± 0.12	0.77 ± 0.11
FS	0.70 ± 0.06	0.63 ± 0.07	0.60 ± 0.11	0.66 ± 0.11

Table 3: NC vs PD classification performances for network features (NF), Voxel Based Morphometry (VBM) and FreeSurfer (FS) obtained with and without the combination of clinical features (CF). Area under the receiver operating characteristics (AUC), accuracy (ACC), sensitivity (sens) and specificity (spec) are reported with the relative standard deviations. Best performance (bold) is obtained with a combined use of network and clinical features (NF+CF). The combination with the clinical features improves the performances in all three cases (NF+CF,VBM+CF,FS+CF).

4.3. Regions of interest

The high discriminative power shown by the features and evaluated in the previous section demonstrates the reliability of the complex network measures as PD markers. Besides the diagnostic support that these features can provide, it is interesting to evaluate which regions result to be affected by the disease and eventually rank them according to their statistical significance.

For each cross-validation round we recorded which features were selected as the most important for classification and outlined the brain regions related to them. Accordingly, for each cross-validation round we counted whether or not a particular anatomic district had been selected and tested the hypothesis that these occurrences had happened by chance. The null hypothesis is that a region should be selected according to a Bernoulli distribution with an *a priori* probability 1/12219, which is the total number of patches. We tested the occurrence of a specific region against the random occurrence rate which was in our case equal to 1/patch number (1/12219), which is in conclusion the number of the carried tests. Figure 6 shows some of the significant regions associated to the diagnosis.

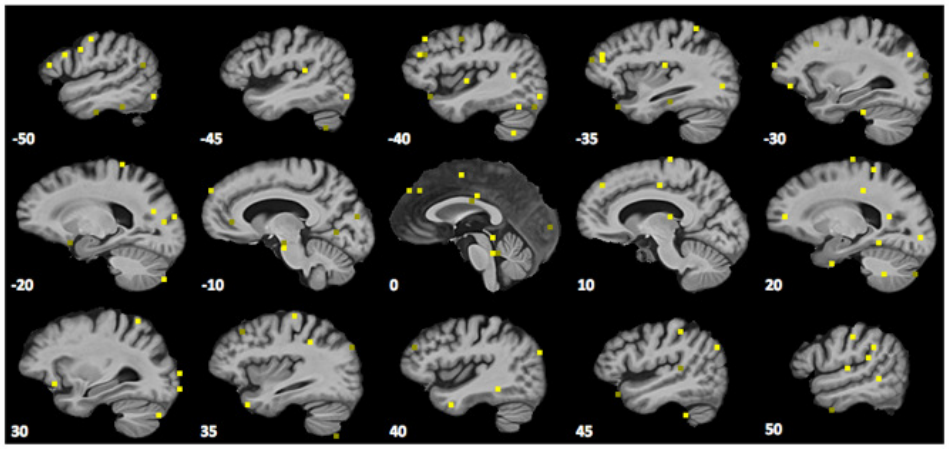

Figure 6: A qualitative overview of significant PD-related patches, as they are outlined by complex network measures, is represented along sagittal planes. More significant patches have a lighter shade of yellow and little by little less significant patches have a darker shade of yellow. The negative planes belong to the left hemisphere while the positive planes belong to the right one.

According to the proposed method, 186 significant patches were detected: 43% within the right hemisphere and 40% within the left one; 9% of patches were located in the cerebellum; brainstem regions appeared in 3% of cases. There is not a great difference between GM and WM regions, in fact the GM regions selected are 53%, the WM ones 46%. The vast majority of brain regions affected by PD lies in the Frontal (27%), Occipital (21%) and Temporal (15%) lobes. For a complete overview of the selected regions and the relative p-values please refer to the Appendix. Instead, we provide in the following Table 4 a compact overview of the first 15 selected patches.

Region	**p-value**
(L) Temporal Lobe. Middle Temporal Gyrus. GM-WM. Ba 39.	$7.2 \cdot 10^{-7}$
(R) Temporal Lobe. Superior and Inferior Temporal Gyrus. GM-WM. Ba 22.	$7.2 \cdot 10^{-7}$
(R) Occipital Lobe. Sub-Gyral (WM).	$3.6 \cdot 10^{-6}$
(L) Occipital Lobe. Superior Occipital Gyrus GM-WM. Ba 19.	$4.6 \cdot 10^{-6}$
(R) Frontal Lobe. Middle Frontal Gyrus (WM).	$5.3 \cdot 10^{-6}$
(L-c) Anterior Lobe. Culmen (GM).	$5.9 \cdot 10^{-6}$
(R) Frontal Lobe. Medial Frontal Gyrus (WM).	$5.9 \cdot 10^{-6}$
(R) Frontal Lobe. Precentral Gyrus. GM. Ba 44.	$5.9 \cdot 10^{-6}$
(L) Limbic Lobe. Cingulate Gyrus. GM. Ba 24.	$5.9 \cdot 10^{-6}$
(R) Parietal Lobe. Precuneus (WM).	$5.9 \cdot 10^{-6}$
(L) Frontal Lobe. Middle Frontal Gyrus. GM. Ba 46.	$6.2 \cdot 10^{-6}$
(L) Brainstem. Midbrain	$6.6 \cdot 10^{-6}$
(L) Temporal Lobe. Fusiform Gyrus GM-WM. Ba 37.	$7.2 \cdot 10^{-6}$
(R-c) Posterior Lobe. Declive. GM.	$7.2 \cdot 10^{-6}$
(R) Temporal Lobe.Fusiform Gyrus. WM.	$7.2 \cdot 10^{-6}$

Table 4: The regions selected according to complex network measures and the inherent level of significance with respect of diagnosis. (L) and (R) denotes the left and right hemispheres; cerebellum regions are denoted with c. Brodmann areas (Ba) are also outlined when appropriate.

These regions have been already detected in several PD studies (Li et al., 2017; Warmuth-Metz et al., 2001; Kim et al., 2013; Wen et al., 2015), another indirect validation of the proposed methodology in that the selected regions consistently correspond to regions whose relationship with the disease is established.

4.4. Robustness of the method

In this section we demonstrate that the proposed complex network description does not require any fine tuning, accordingly we used for both image processing and classification standard configurations, then we evaluated, a

posteriori, that choosing other configurations would not have affected the obtained results. We evaluated the relation between the threshold used to remove some edges from the network and the classification accuracy; a wide range of thresholds was explored, see Figure 7. Threshold was varied from 0 to 0.9 with 0.1 steps and with a fixed patch volume of 125 voxels. The maximum value of the classification accuracy was obtained at 0.3 threshold. It is worth noting that this value also corresponded to minimum variance. Accuracy remained constant at 0.93 for a wide range of correlations $[0.3, 0.5]$, thus confirming that the method does not require a fine tuning of threshold values. For higher threshold values, there was a significant performance drop, suggesting that too high threshold values cause the loss of important links.

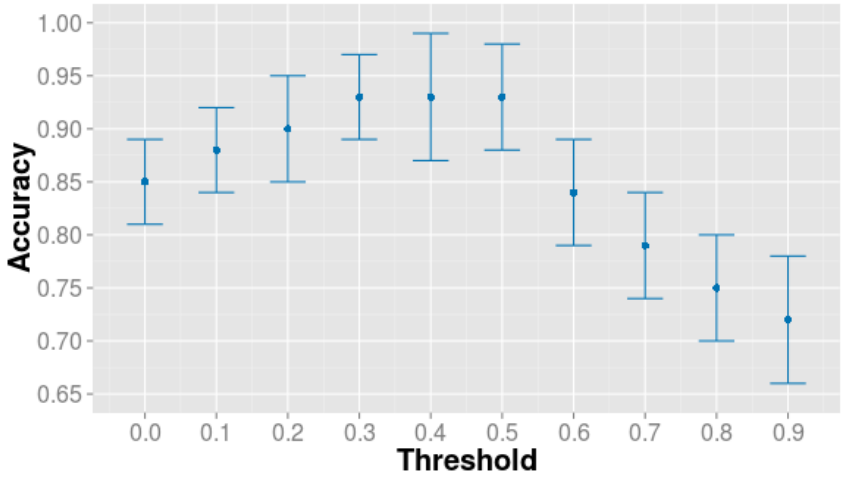

Figure 7: The figure shows the accuracy as a function of the threshold that changes from 0 to 0.9. In correspondence of a threshold value of 0.3, the best accuracy and the minimal standard deviation were reached.

As demonstrated in section 4.1, the proposed complex network approach significantly enhances the discriminative power of clinical features. The network measures derived from MRI data effectively characterize PD patterns. To evaluate the robustness of the informative content provided by our method, we explored the hyperparameter space. Firstly, we evaluated the cost parameter which plays a fundamental role for Support Vector Machines, see Figure 8.

In fact, the cost determines how much the SVM model should fit the training data by varying the margins of the decision hyperplane, larger values

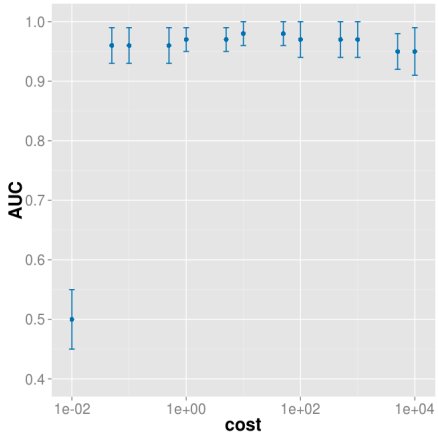

Figure 8: The classification performance in terms of AUC remains stable by varying the cost parameter. For tiny cost values the decision hyperplane margins are too large and the performance drops.

of cost correspond to smaller margins. The results show that for a wide range of cost values the classification performance remains stable granting robust results. When the cost reaches the 0.01 value the performance drops, this means that the margins have become so large that the model cannot just fit the data. Moreover, we investigated the model robustness with respect of the gamma parameter which defines how far the region of influence of each training example should extend, see Figure 9.

As gamma controls the variance of the model, by varying gamma one can move from a high-bias to a high-variance model. Of course, the optimal classification region stands between these two cases. The results show that for the present model a wide stability region exists, in fact the AUC consistently remains over the 0.90 value for a gamma variation of more than 4 orders of magnitude.

4.5. Evaluation of the informative content

To evaluate the goodness of complex network measures as PD markers it is not sufficient to demonstrate that these features allow an accurate and robust classification. First of all, looking at the classification performance it is not possible to guess if credit should be given to the SVM classification model or the informative content of the features. As such, we compared the classification performance of several state-of-the-art classifiers, specifically we investigated Random Forest (RF), Naive Bayes (NB) and Neural

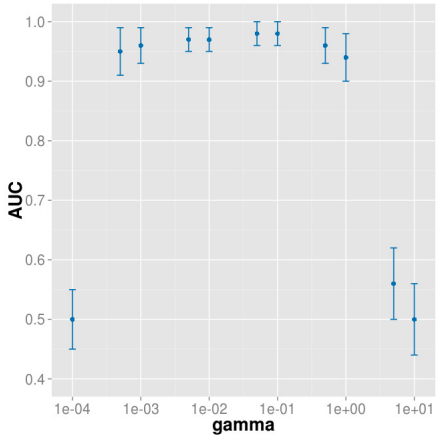

Figure 9: Varying the gamma parameter it is possible to switch from high-bias to high-variance models, in this case however the optimal classification region extends for more than 4 orders of magnitude.

Network (NN) classifiers. For each method we explored within a nested cross-validation the hyperparameter space and several configurations, only optimal configuration results are reported in the following Table 5.

method	**AUC**	**ACC**	**sens**	**spec**
Neural Network	0.94 ± 0.04	0.89 ± 0.05	0.90 ± 0.08	0.88 ± 0.07
Random Forest	0.97 ± 0.02	0.91 ± 0.05	0.90 ± 0.07	0.91 ± 0.07
Naive Bayes	0.97 ± 0.03	0.92 ± 0.05	0.91 ± 0.07	$\mathbf{0.93 \pm 0.07}$
Support Vector Machine	$\mathbf{0.97 \pm 0.02}$	$\mathbf{0.93 \pm 0.04}$	$\mathbf{0.93 \pm 0.06}$	0.92 ± 0.07

Table 5: A comparison between different machine learning methods (Neural Networks, Random Forests, Naive Bayes and Support Vector Machine classifiers) shows that the proposed complex network approach allows a robust diagnosis independently from the choice of the classifier, although Support Vector Machine reaches slightly better results (in bold) for almost each metric: area under the receiving operating characteristics (AUC), accuracy (ACC), sensitivity (sens) and specificity (spec).

The table shows that no significant difference can be found between different models, even if SVM would seem to perform slightly better than the others. These results demonstrate that beyond the differences due to the machine learning models adopted, the proposed approach yields an outstanding base of knowledge for PD discrimination.

However, this test does not evaluate the agreement between the models. In fact, in principle two distinct models could perform equally but misclassi-

fying different subjects. In order to assess the agreement between our chosen SVM model and the other models, we investigated the relationships existing between the classification scores. The results are presented in Figure 10.

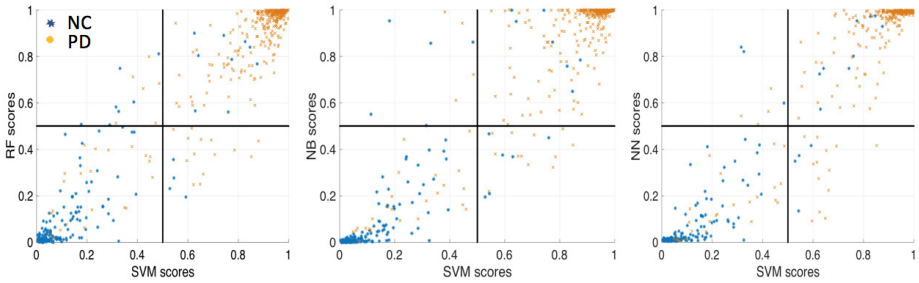

Figure 10: From left to right the agreement between the scores obtained with Support Vector Machine (SVM) and those obtained by: Random Forest (RF), Naive Bayes (NB) and Neural Network (NN) classifiers. The scores are densely distributed in top right and bottom left quadrants, where their predictions agree. Looking at the top left and bottom right quadrants it can be noted that, when in disagreement, SVM scores tend to be slightly more accurate than other scores.

The classification scores are densely distributed in the top right and bottom left quadrants. The top right quadrant includes subjects whose classification scores exceed the 0.5 value, it is the case of subjects diagnosed with PD from both the SVM model, which is always reported on the x axis, and the other models, which are reported instead on the y axis. Analogously, the bottom left quadrant includes those subjects, whose classification scores are lower than 0.5, for which the models agree assigning a NC status. As expected from previous measures, the models correctly distinguish the two classes, in fact in the top right quadrant the vast majority of subjects is shown in orange, as subjects have mostly a PD diagnosis, and in the bottom left the vast majority is in blue, as subjects are mainly NC.

The top left and bottom right quadrants are the regions of disagreement. In these regions, in fact, the SVM model assigns a diagnosis different from other models. For example, a subject belonging to the bottom right quadrant has an SVM score > 0.5 and it is accordingly diagnosed as PD but a RF score (or NB/NN) < 0.5 and it is therefore labeled as NC. First of all, it is worth noting that these two quadrants are sparsely populated, especially compared to the top right and bottom left ones, therefore this is a further demonstration of the agreement between the models; besides, SVM tends to be more accurate.

Keeping on with our example, in the bottom right quadrant for all the three cases the majority of subjects is orange, meaning that their true label is PD. This means that the SVM predictions is the right one. The same consideration holds for the top left quadrant, where the majority of subjects is represented in blue, and, again SVM correctly labels them as NC.

4.6. Scale study and VBM

The proposed approach depends on the size of the brain patches used for the complex network model. In our previous studies concerning Alzheimer's disease (La Rocca et al., 2017; Amoroso et al., 2017) we observed that one key aspect of complex network descriptions is that they let naturally emerge a dimensional scale, which is typical of the disease. For example, for Alzheimer characterization the best results were obtained with patches having approximately a volume of 3000 mm^3. Accordingly, in this work we explored a wide range of patch volumes and measured the training classification accuracy, see Figure 11.

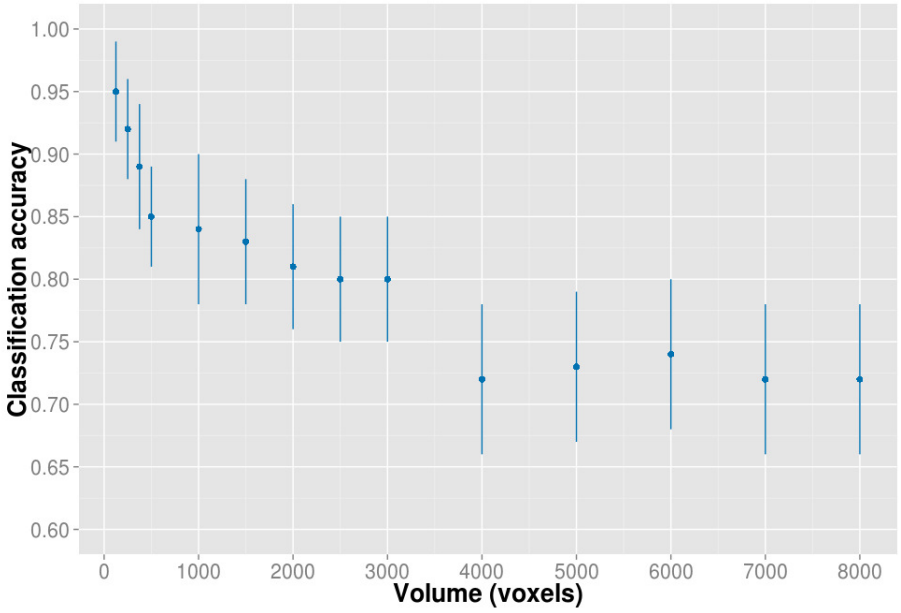

Figure 11: The optimal patch dimension expresses the existence of a preferred dimension or scale for PD markers. In particular, best classification accuracy 0.95 ± 0.04 is obtained with smaller patches ($5 \times 5 \times 5$ voxels). For larger patch dimensions the performance drops and reaches a stable plateau.

The classification accuracy decreases monotonically. The best performance was obtained with patches of $5 \times 5 \times 5$ voxels. This result is significantly different from what we observed in Alzheimer. Moreover, when the patch volume reaches 4000 voxels (we remind here that for the present study voxels and mm^3 can be interchangeably used) the accuracy remains constantly around 0.73.

We used a standard VBM pipeline to segment gray and white matter of each MRI scan. Then we normalized each subject to the MNI152 template and extracted the t-Student maps to determine if some clusters of voxels (> 30) exhibit an association with the diagnosis. We found a good agreement with the regions detected with our complex network description, see Figure 12 for an overview, but, remarkably, the number of regions showing an association with the clinic was consistently reduced.

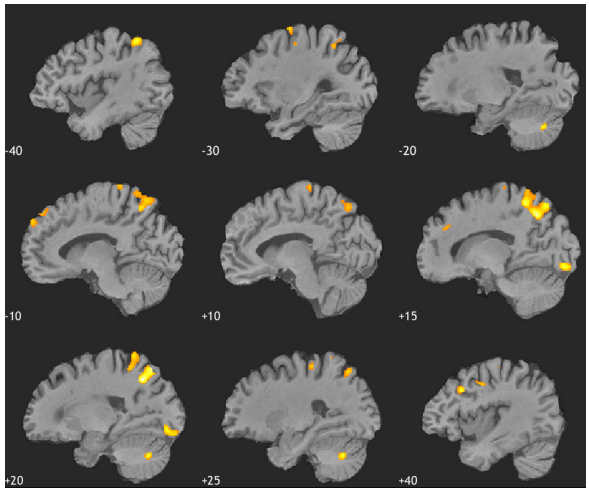

Figure 12: Voxel based morphometry shows the presence of some clusters (the sagittal plane is reported), however these regions represent only a subset of those outlined by our approach.

However, as explained in Section 1, voxel-wise approaches have an intrinsic drawback in that the need for thousands of multiple comparisons dramatically lowers the statistical power of commonly available datasets. Indeed, we found no significant association after Bonferroni corrections.

5. Discussion

The PD onset is characterized by clinical symptoms which emerge when the dopaminergic deficit has reached a considerable level. Therapies or drugs could easily be ineffective at this stage. This is why the identification of accurate markers, and hopefully of a diagnostic framework, based on symptoms related to the prodromal or early phases of the disease is urgent. Our approach uses complex network measures to characterize PD patterns and develop a fully-automated machine learning diagnosis support system. The proposed methodology is robust and accurate. In addition, it provides detailed information about the brain regions mostly affected by the disease as it ranks them by associating an easy-to-interpret level of significance; thus, this method opens the possibility for further comprehension of PD patterns.

The proposed approach reaches an accurate diagnosis (AUC $= 0.97 \pm 0.02$ and ACC $= 0.93 \pm 0.04$) and these results compare favorably with other state-of-the-art approaches. Among the most recent methodologies, the joint feature-sample selection algorithm by (Adeli et al., 2016) reports an 82% accuracy and currently achieves one of the best classification performances when using only MRI data. It is worth mentioning that other studies using structural MRI features reported interesting results, such as (Salvatore et al., 2014) whose VBM-based methodology allowed an accuracy of 83.2%, although a significantly smaller sample including only 28 controls and 28 PD subjects.

It has been recently shown that accurate diagnosis (97.5%) can be obtained when combining both MRI and SPECT data (Adeli et al., 2017). However, such a study demonstrates that the classification accuracy almost relies on SPECT as SPECT provides a diagnostic accuracy of 95.6% when used without MRI. Thus, MRI data seems to slightly contribute to diagnosis accuracy. Nevertheless, as SPECT detects the substantial loss of dopaminergic neurons, markers based on this imaging modalities could be better employed in later stages of the disease, for example when motor symptoms appear.

Our work emphasizes the possibility to conveniently use complex network measures as PD markers. We demonstrated that, besides the high accuracy, MRI features based on complex networks bring a significant improvement to classification based only on clinical features. Classification is balanced, in fact specificity and sensitivity give similar results, unlike what we observed with clinical features which tend to be more specific but less sensitive. This

can be expected as the subjects included in this study are all considered at the baseline when the clinical symptoms are still mild.

Previous studies have usually investigated PD patterns basing on a restricted list of regions of interest (Braak et al., 2003; Burke et al., 2008). The reasons for such a choice are twofold: firstly, from a clinical perspective it is known that some regions are affected by the disease only at its later stages, as a consequence these regions can be safely disregarded; secondly, whole brain analysis can be too computationally intensive and when data samples are small it is easy to lack the statistical power required to detect small effects.

On the other hand, focusing on some regions can prevent the detection of interesting effects in the brain regions excluded or decrease the discrimination power of the approach. For example, (Worker et al., 2014) found no significant cortical changes between PD patients and controls when examining a restricted list of brain regions.

It is worth noting that ROI-based approaches are intrinsically biased by segmentation errors which can prevent the methodologies to reach optimal sensitivity. This is why, even if using an ROI approach, we preferred a whole brain description. Our complex network approach significantly reduced the computational burden yielded by voxel-based approaches even if the number of examined regions was higher than in ROI-based studies.

In addition, our method is more sensitive than a standard VBM; as reported in previous studies, see for example (Focke et al., 2011), with VBM we observed no significant changes in cortical morphology when comparing NC and PD subjects. On the other hand the regions outlined with the proposed methodology are consistent with previous studies. Temporal and Frontal Gyri changes have shown atrophic patterns, especially in patients with dementia (Burton et al., 2004; Xia et al., 2013); cognitive impairment seems to acquire a relevant role for diagnosis also because of the inclusion of Brodmann areas $24, 37, 44$ and 46 as shown also in (Burton et al., 2004; Hughes et al., 1992; Nagano-Saito et al., 2005). As expected, Substantia Nigra and brain midstem also play a relevant role for the diagnosis. We found that most of the significant regions are not adjacent, indeed there is no a priori reason why adjacent patches (which can often include distinct anatomical districts) should share the same informative content, as neurodegenerative diseases may have a diffuse effect that involves multiple voxels not necessarily belonging to the same anatomical region Burton et al. (2004).

Our results outline the important role of combining MRI and clinical

features for an accurate early diagnosis. In fact, besides the increment of the classification accuracy, it is manifest that the use of clinical features is biased towards the NC class. We demonstrated that classification scores based only on clinical features were poorly sensitive, with lots of PD patients misclassified as controls; this effect is reasonable as in the early phase of the disease clinical symptoms are mild. On the other hand this result outlines the importance of complex network markers to improve both sensitivity and specificity of the classification.

6. Conclusions

In this work, we have demonstrated how complex networks can proficiently be used to define a novel brain connectivity and consequently introduce accurate markers for PD. We evaluated the robustness and the accuracy of the proposed methodology with both a direct evaluation, involving the measure of classification metrics, and an indirect check, regarding the brain regions mostly affected by the disease. We validated our method on a mixed cohort of controls and patients from the PPMI dataset; the proposed methodology compares well with other state-of-the-art approaches for what concerns NC/PD classification. In addition, our method allowed an investigation of the brain regions related to the disease starting from a segmentation completely unsupervised over the whole brain without the necessity to *a priori* focus on specific anatomical regions, a fundamental aspect when looking for novel markers. Our results confirm what has been found in other studies and outlines new interesting aspects, specifically: (i) our work demonstrates that MRI data, and in particular complex network measures, provide an efficient and accurate description of PD patterns; (ii) novel MRI markers combined with clinical scores typical of prodromal PD can be used for an accurate early diagnosis; this approach (iii) compares favorably with state-of-the-art methodologies basing on MRI data and (iv) compares well with methodologies including other imaging modalities such as SPECT. In brief, our work shows that the connectivity of several brain regions is significantly related to PD. Thus, we hope this result will stimulate further investigations to better understand the disease and its mechanisms. These results also suggest the applicability of the methodology to support PD diagnosis in clinical practice and possibly other disease affecting brain connectivity. Further studies could investigate how to improve this methodology, for example using multi-modal imaging data. In addition, it would be interesting to provide a comprehen-

sive model for the regions outlined by our approach from a more specifically clinical perspective.

Acknowledgments

PPMI, a public-private partnership,is funded by the Michael J. Fox Foundation for Parkinson's Research and other funding partners include AbbVie, Avid Radiopharmaceuticals, Biogen Idec, Bristol-Myers Squibb, Covance, GE Healthcare, Genentech, GlaxoSmithKline, Eli Lilly and Company, Lundbeck, Merck & Co., Meso Scale Discovery, Pfizer, Piramal, Hoffmann-La Roche, and UCB (Union ChimiqueBelge). All authors disclose any actual or potential conflicts of interest, including any financial, personal, or other relationships with other people or organizations that could inappropriately influence their work. All experiments were performed with the informed consent of each participant or caregiver in line with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Local institutional ethics committees approved the study.

References

Adeli, E., Shi, F., An, L., Wee, C.-Y., Wu, G., Wang, T., Shen, D., 2016. Joint feature-sample selection and robust diagnosis of Parkinson's disease from MRI data. NeuroImage 141, 206–219.

Adeli, E., Wu, G., Saghafi, B., An, L., Shi, F., Shen, D., 2017. Kernel-based Joint Feature Selection and Max-Margin Classification for Early Diagnosis of Parkinson's Disease. Scientific Reports 7, 41069.

Amoroso, N., Bellotti, R., Diacono, D., La Rocca, M., Tangaro, S., 2017. Salient Networks: A Novel Application to Study Brain Connectivity. In: International Conference on Bioinformatics and Biomedical Engineering. Springer, pp. 444–453.

Antonini, A., Leenders, K. L., Vontobel, P., Maguire, R. P., Missimer, J., Psylla, M., Günther, I., 1997. Complementary PET studies of striatal neuronal function in the differential diagnosis between multiple system atrophy and Parkinson's disease. Brain 120 (12), 2187–2195.

Ashburner, J., Friston, K. J., 2000. Voxel-based morphometrythe methods. Neuroimage 11 (6), 805–821.

Beyer, M. K., Janvin, C. C., Larsen, J. P., Aarsland, D., 2007. A magnetic resonance imaging study of patients with Parkinson's disease with mild cognitive impairment and dementia using voxel-based morphometry. Journal of Neurology, Neurosurgery & Psychiatry 78 (3), 254–259.

Bouwmans, A. E., Vlaar, A. M., Mess, W. H., Kessels, A., Weber, W. E., 2013. Specificity and sensitivity of transcranial sonography of the substantia nigra in the diagnosis of Parkinson's disease: prospective cohort study in 196 patients. BMJ open 3 (4), e002613.

Braak, H., Del Tredici, K., Rüb, U., de Vos, R. A., Steur, E. N. J., Braak, E., 2003. Staging of brain pathology related to sporadic Parkinson's disease. Neurobiology of aging 24 (2), 197–211.

Breiman, L., 1996. Bagging predictors. Machine learning 24 (2), 123–140.

Breiman, L., 2001. Random forests. Machine learning 45 (1), 5–32.

Bullmore, E., Sporns, O., 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience 10 (3), 186–198.

Burke, R. E., Dauer, W. T., Vonsattel, J. P. G., 2008. A critical evaluation of the Braak staging scheme for Parkinson's disease. Annals of neurology 64 (5), 485–491.

Burton, E. J., McKeith, I. G., Burn, D. J., Williams, E. D., OBrien, J. T., 2004. Cerebral atrophy in Parkinson's disease with and without dementia: a comparison with Alzheimer's disease, dementia with Lewy bodies and controls. Brain 127 (4), 791–800.

Chagas, M. H. N., Tumas, V., Pena-Pereira, M. A., Machado-de Sousa, J. P., dos Santos, A. C., Sanches, R. F., Hallak, J. E., Crippa, J. A. S., 2017. Neuroimaging of major depression in Parkinson's disease: Cortical thickness, cortical and subcortical volume, and spectroscopy findings. Journal of Psychiatric Research 90, 40–45.

Chaudhuri, K. R., Healy, D. G., Schapira, A. H., 2006. Non-motor symptoms of Parkinson's disease: diagnosis and management. The Lancet Neurology 5 (3), 235–245.

Cherubini, A., Morelli, M., Nisticó, R., Salsone, M., Arabia, G., Vasta, R., Augimeri, A., Caligiuri, M. E., Quattrone, A., 2014. Magnetic resonance support vector machine discriminates between Parkinson disease and progressive supranuclear palsy. Movement Disorders 29 (2), 266–269.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine learning 20 (3), 273–297.

Daianu, M., Jahanshad, N., Villalon-Reina, J. E., Mendez, M. F., Bartzokis, G., Jimenez, E. E., Joshi, A., Barsuglia, J., Thompson, P. M., 2014. Rich club network analysis shows distinct patterns of disruption in frontotemporal dementia and Alzheimer's disease. In: Computational Diffusion MRI. Springer, pp. 13–22.

Dalrymple-Alford, J., MacAskill, M., Nakas, C., Livingston, L., Graham, C., Crucian, G., Melzer, T., Kirwan, J., Keenan, R., Wells, S., et al., 2010. The MoCA well-suited screen for cognitive impairment in Parkinson disease. Neurology 75 (19), 1717–1725.

Díaz-Uriarte, R., De Andres, S. A., 2006. Gene selection and classification of microarray data using random forest. BMC bioinformatics 7 (1), 3.

Dorsey, E., Constantinescu, R., Thompson, J., Biglan, K., Holloway, R., Kieburtz, K., Marshall, F., Ravina, B., Schifitto, G., Siderowf, A., et al., 2007. Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. Neurology 68 (5), 384–386.

Duchesne, S., Rolland, Y., Vérin, M., 2009. Automated computer differential classification in Parkinsonian syndromes via pattern analysis on MRI. Academic radiology 16 (1), 61–70.

Fischl, B., 2012. Freesurfer. Neuroimage 62 (2), 774–781.

Focke, N. K., Helms, G., Scheewe, S., Pantel, P. M., Bachmann, C. G., Dechent, P., Ebentheuer, J., Mohr, A., Paulus, W., Trenkwalder, C., 2011. Individual voxel-based subtype prediction can differentiate progressive supranuclear palsy from idiopathic Parkinson syndrome and healthy controls. Human brain mapping 32 (11), 1905–1915.

Friedman, J., Friedman, H., 1993. Fatigue in Parkinson's disease. Neurology 43 (10), 2016–2016.

765 Gagnon, J.-F., Bédard, M.-A., Fantini, M., Petit, D., Panisset, M., Rompre,
766 S., Carrier, J., Montplaisir, J., 2002. REM sleep behavior disorder and
767 REM sleep without atonia in Parkinson's disease. Neurology 59 (4), 585–
768 589.

769 Gibb, W., Lees, A., 1988. The relevance of the lewy body to the pathogenesis
770 of idiopathic parkinson's disease. Journal of Neurology, Neurosurgery &
771 Psychiatry 51 (6), 745–752.

772 Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S.,
773 Martinez-Martin, P., Poewe, W., Sampaio, C., Stern, M. B., Dodel, R.,
774 et al., 2008. Movement Disorder Society-sponsored revision of the Unified
775 Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and
776 clinimetric testing results. Movement disorders 23 (15), 2129–2170.

777 Hansen, A. K., Knudsen, K., Lillethorup, T. P., Landau, A. M., Parbo, P.,
778 Fedorova, T., Audrain, H., Bender, D., Østergaard, K., Brooks, D. J.,
779 et al., 2016. In vivo imaging of neuromelanin in Parkinson's disease using
780 18F-AV-1451 PET. Brain, aww098.

781 Hirschauer, T. J., Adeli, H., Buford, J. A., 2015. Computer-aided diagnosis of
782 Parkinson's disease using enhanced probabilistic neural network. Journal
783 of medical systems 39 (11), 179.

784 Hoehn, M. M., Yahr, M. D., et al., 1998. Parkinsonism: onset, progression,
785 and mortality. Neurology 50 (2), 318–318.

786 Huber, S. J., Shuttleworth, E. C., Paulson, G. W., 1986. Dementia in Parkin-
787 son's disease. Archives of Neurology 43 (10), 987–990.

788 Hughes, A. J., Daniel, S. E., Kilford, L., Lees, A. J., 1992. Accuracy of clinical
789 diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of
790 100 cases. Journal of Neurology, Neurosurgery & Psychiatry 55 (3), 181–
791 184.

792 Jankovic, J., 2008. Parkinson's disease: clinical features and diagnosis. Jour-
793 nal of Neurology, Neurosurgery & Psychiatry 79 (4), 368–376.

794 Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimiza-
795 tion for the robust and accurate linear registration and motion correction
796 of brain images. Neuroimage 17 (2), 825–841.

Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Medical image analysis 5 (2), 143–156.

Johns, M. W., et al., 1991. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. sleep 14 (6), 540–545.

Kalia, L. V., Lang, A. E., April 2015. Parkinson's disease. Lancet 386 (9996), 896–912.

Kim, H. J., Kim, S. J., Kim, H. S., Choi, C. G., Kim, N., Han, S., Jang, E. H., Chung, S. J., Lee, C. S., 2013. Alterations of mean diffusivity in brain white matter and deep gray matter in parkinson's disease. Neuroscience letters 550, 64–68.

Kwon, D.-H., Kim, J.-M., Oh, S.-H., Jeong, H.-J., Park, S.-Y., Oh, E.-S., Chi, J.-G., Kim, Y.-B., Jeon, B. S., Cho, Z.-H., 2012. Seven-tesla magnetic resonance images of the substantia nigra in Parkinson disease. Annals of neurology 71 (2), 267–277.

La Rocca, M., Amoroso, N., Bellotti, R., Diacono, D., Monaco, A., Monda, A., Tateo, A., Tangaro, S., 2017. A Multiplex Network Model to Characterize Brain Atrophy in Structural MRI. In: Emergent Complexity from Nonlinearity, in Physics, Engineering and the Life Sciences. Springer, pp. 189–198.

Li, X., Xing, Y., Schwarz, S. T., Auer, D. P., 2017. Limbic grey matter changes in early parkinson's disease. Human Brain Mapping.

Litvan, I., Goldman, J. G., Tröster, A. I., Schmand, B. A., Weintraub, D., Petersen, R. C., Mollenhauer, B., Adler, C. H., Marder, K., Williams-Gray, C. H., et al., 2012. Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society Task Force guidelines. Movement Disorders 27 (3), 349–356.

Lo, C.-Y., Wang, P.-N., Chou, K.-H., Wang, J., He, Y., Lin, C.-P., 2010. Diffusion tensor tractography reveals abnormal topological organization in structural cortical networks in Alzheimer's disease. Journal of Neuroscience 30 (50), 16876–16885.

Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., Coffey, C., Kieburtz, K., Flagg, E., Chowdhury, S., et al., 2011. The parkinson

progression marker initiative PPMI. Progress in neurobiology 95 (4), 629–635.

Marquand, A. F., Filippone, M., Ashburner, J., Girolami, M., Mourao-Miranda, J., Barker, G. J., Williams, S. C., Leigh, P. N., Blain, C. R., 2013. Automated, high accuracy classification of parkinsonian disorders: a pattern recognition approach. PloS one 8 (7), e69237.

Masdeu, J. C., 2017. Future Directions in Imaging Neurodegeneration. Current Neurology and Neuroscience Reports 17 (1), 9.

Menichetti, G., Remondini, D., Panzarasa, P., Mondragón, R. J., Bianconi, G., 2014. Weighted multiplex networks. PloS one 9 (6), e97857.

Miller, D. B., O'Callaghan, J. P., 2015. Biomarkers of Parkinson's disease: present and future. Metabolism 64 (3), S40–S46.

Nagano-Saito, A., Washimi, Y., Arahata, Y., Kachi, T., Lerch, J., Evans, A., Dagher, A., Ito, K., 2005. Cerebral atrophy and its relation to cognitive impairment in Parkinson disease. Neurology 64 (2), 224–229.

Pilotto, A., Yilmaz, R., Berg, D., 2015. Developments in the role of transcranial sonography for the differential diagnosis of parkinsonism. Current neurology and neuroscience reports 15 (7), 1–10.

Postuma, R. B., Aarsland, D., Barone, P., Burn, D. J., Hawkes, C. H., Oertel, W., Ziemssen, T., 2012. Identifying prodromal Parkinson's disease: Pre-Motor disorders in Parkinson's disease. Movement Disorders 27 (5), 617–626.

Postuma, R. B., Berg, D., Stern, M., Poewe, W., Olanow, C. W., Oertel, W., Obeso, J., Marek, K., Litvan, I., Lang, A. E., et al., 2015. MDS clinical diagnostic criteria for Parkinson's disease. Movement Disorders 30 (12), 1591–1601.

Saeys, Y., Inza, I., Larrañaga, P., 2007. A review of feature selection techniques in bioinformatics. bioinformatics 23 (19), 2507–2517.

Salvatore, C., Cerasa, A., Castiglioni, I., Gallivanone, F., Augimeri, A., Lopez, M., Arabia, G., Morelli, M., Gilardi, M., Quattrone, A., 2014.

Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and Progressive Supranuclear Palsy. Journal of Neuroscience Methods 222, 230–237.

Singaram, C., Gaumnitz, E., Torbey, C., Ashraf, W., Quigley, E., Sengupta, A., Pfeiffer, R., 1995. Dopaminergic defect of enteric nervous system in Parkinson's disease patients with chronic constipation. The Lancet 346 (8979), 861–864.

Singh, G., Samavedham, L., 2015. Unsupervised learning based feature extraction for differential diagnosis of neurodegenerative diseases: a case study on early-stage diagnosis of Parkinson disease. Journal of neuroscience methods 256, 30–40.

Smith, S. M., 2002. Fast robust automated brain extraction. Human brain mapping 17 (3), 143–155.

Stam, C., Jones, B., Nolte, G., Breakspear, M., Scheltens, P., 2007. Small-world networks and functional connectivity in Alzheimer's disease. Cerebral cortex 17 (1), 92–99.

Stiasny-Kolster, K., Mayer, G., Schäfer, S., Möller, J. C., Heinzel-Gutenbrunner, M., Oertel, W. H., 2007. The REM sleep behavior disorder screening questionnairea new diagnostic instrument. Movement disorders 22 (16), 2386–2393.

Summerfield, C., Junqué, C., Tolosa, E., Salgado-Pineda, P., Gómez-Ansón, B., Martí, M. J., Pastor, P., Ramírez-Ruíz, B., Mercader, J., 2005. Structural brain changes in Parkinson disease with dementia: a voxel-based morphometry study. Archives of Neurology 62 (2), 281–285.

Suwijn, S. R., van Boheemen, C. J., de Haan, R. J., Tissingh, G., Booij, J., de Bie, R. M., 2015. The diagnostic accuracy of dopamine transporter SPECT imaging to detect nigrostriatal cell loss in patients with Parkinson's disease or clinically uncertain parkinsonism: a systematic review. EJNMMI research 5 (1), 12.

Tessa, C., Lucetti, C., Giannelli, M., Diciotti, S., Poletti, M., Danti, S., Baldacci, F., Vignali, C., Bonuccelli, U., Mascalchi, M., et al., 2014. Progression of brain atrophy in the early stages of Parkinson's disease: A

longitudinal tensor-based morphometry study in de novo patients without cognitive impairment. Human brain mapping 35 (8), 3932–3944.

Tijms, B. M., Wink, A. M., de Haan, W., van der Flier, W. M., Stam, C. J., Scheltens, P., Barkhof, F., 2013. Alzheimer's disease: connecting findings from graph theoretical studies of brain networks. Neurobiology of aging 34 (8), 2023–2036.

Warmuth-Metz, M., Naumann, M., Csoti, I., Solymosi, L., 2001. Measurement of the midbrain diameter on routine magnetic resonance imaging: a simple and accurate method of differentiating between parkinson disease and progressive supranuclear palsy. Archives of neurology 58 (7), 1076–1079.

Wen, M.-C., Ng, A., Chander, R. J., Au, W. L., Tan, L. C., Kandiah, N., 2015. Longitudinal brain volumetric changes and their predictive effects on cognition among cognitively asymptomatic patients with parkinson's disease. Parkinsonism & related disorders 21 (5), 483–488.

Worker, A., Blain, C., Jarosz, J., Chaudhuri, K. R., Barker, G. J., Williams, S. C., Brown, R., Leigh, P. N., Simmons, A., 2014. Cortical thickness, surface area and volume measures in Parkinson's disease, multiple system atrophy and progressive supranuclear palsy. PloS one 9 (12), e114167.

Xia, J., Miu, J., Ding, H., Wang, X., Chen, H., Wang, J., Wu, J., Zhao, J., Huang, H., Tian, W., et al., 2013. Changes of brain gray matter structure in Parkinson's disease patients with dementia. Neural regeneration research 8 (14), 1276.

Yesavage, J., Brink, T., Rose, T., et al., 2000. Geriatric depression scale (GDS). Handbook of psychiatric measures. Washington DC: American Psychiatric Association, 544–6.

Yesavage, J. A., Sheikh, J. I., 1986. 9/Geriatric depression scale (GDS) recent evidence and development of a shorter version. Clinical gerontologist 5 (1-2), 165–173.