Automatic classification of idiopathic Parkinsonian disease and progressive supranuclear palsy using multispectral MRI datasets

Sahand Talai

Supervisor: Nils Daniel Forkert, PhD

I. Introduction

Parkinson's disease (PD), the second most common progressive neurodegenerative disease after Alzheimer's disease (AD), is characterized by a range of motor and non-motor symptoms. Predominant motor symptoms include, resting tremor, rigidity, postural instability and bradykinesia, whereas, olfactory dysfunction, sleep disorder, autonomic and cognitive dysfunction are the main non-motor related symptoms[1]. PD can be categorized into idiopathic Parkinsonian disease (IPD) and atypical Parkinsonian syndromes, which includs progressive supranuclear palsy (PSP), multiple system atrophy (MSA), and corticobasal degeneration (CBD). The underlying physiological causes of IPD is still debated, however, it is widely accepted that it is mainly associated with a progressive neuronal loss in pars compacta of substantia nigra (SN) and also in other pigmented nuclei[2] followed by white matter (WM) vitiation[3,4]. Moreover, several studies have shown a slight reduction in cortical thickness and volume in IPD compared to healthy controls (HC)[5]. In PSP, midbrain atrophy along with the widening of the third ventricular and tegmental atrophy are observed[6]. In midbrain atrophy anteroposterior midbrain diameter is reduced and an abnormal superior profile which consists of a flat or concave compared to a convex in healthy controls is observed[5]. The predominant neurological changes in PSP consists of neuronal loss and gliosis in the globus pallidus and substantia nigra[7]. MSA can be classified into MSA-p and MSA-c based on the predominant features of Parkinsonian and cerebellar ataxia, respectively[8]. In MSA, a significant reduction in mean striatal, brainstem and cerebellar volumes along with neuronal loss and gliosis in the putamen and substantia nigra are observed[5]. MSA-p is recognized by massive degeneration of the nigrostriatal system, whereas in MSA-c the olivopontocerebellar system is affected[9,10]. In CBD, the deposition of tau proteins in neurons inside the thalamus has been observed in postmortem studies[11]. Moreover, asymmetric cortical atrophy and putaminal hypo intensity are present in this form of the disease[12]. As of now, there are no clinical procedures to stop PD progression. However, some common therapies to slow down disease progression or mitigate the symptoms include medication, cell transplantation, and deep brain stimulation[13].

An estimated seven to ten million people worldwide are living with parkinson's disease[14]. 55,000 Canadians aged 18 or older are reported to have Parkinson's disease whereas 97% of them are 65 years old or older[15]. In the United States, about one million patients are diagnosed with parkinson's disease whereas 60,000 are being diagnosed annually[14]. Overall, statistical data suggests that men are one and a half times more likely to have PD than women. The combined direct and indirect costs of PD in the US is estimated to be 25 billion dollars, whereas medicinal costs might reach up to 3000 dollars per year[14].

Parkinson's disease diagnosis is commonly conducted by symptom monitoring along with criteria defined by the UK Parkinson's Disease Society Brain Bank[16], the US National Institute of Neurological Disorders and stroke[16] and the Unified Parkinson's Disease Rating Scale[17] (UPDRS). However, the main drawback of these diagnosis methods is that syndrome specific symptoms between IPD and the various atypical entities overlap, especially in the early stages of PD. This might cause inaccurate initial diagnosis, which leads to false medication administration which in turn causes suboptimal disease prognosis. The correct recognition of PD syndromes is of the outmost importance since early stage syndrome specific medicinal intervention has been shown to mitigate symptoms significantly[18]. Moreover, up to 25% of initial diagnoses are falsely declared as IPD[17], which is a significant margin of error. In order to compensate for these inaccuracies, computer aided diagnosis (CAD), such as image based PD classification methods have been proposed to assist clinicians in their diagnosis of Parkinson's disease. These methods utilize functional image based biomarkers obtained from Positron Emission Tomography (PET) and MRI datasets[19,20], or morphological information extracted from high-resolution anatomical MRI[21,22]. Several methods for the image-based classification of IPD and PSP patients have been proposed in the past. Within this context, high-resolution T1-weighted MRI datasets have been used for classifying IPD and PSP patient based on atrophy-related differences in regional brain volumes[22]. Other studies have shown that positron emission tomography[19], as well as T2*[23]and susceptibility-weighted imaging (SWI)[24]MRI datasets are also useful to differentiate IPD and PSP

patients based on differences regarding the regional brain iron metabolism or local brain iron accumulation. Furthermore, another recent study suggests that diffusion tensor MR imaging (DTI), which can be used to analyze and quantify micro structural changes in brain tissue integrity and connectivity, is also a valuable image sequence for an automatic classification of PSP and IPD patients[25]. Each of the MRI modalities described have been used to identify differences between IPD and PSP patients in the past while some of these image-based biomarkers have also been used within automatic classification methods. However, in most cases the reported classification precision was between 80 and 90%, which is still not sufficient for actual clinical applications. Moreover, there have been attempts to use non-image based markers in classifiers to differentiate between PD and healthy controls. These classifiers use biomarkers obtained from voice measurements[17], gait analysis[26], hand writing analysis[27], electromyography (EMG)[28], odor detection tests[29] and blood test[30].

Parkinson's disease biomarkers have been used to train and test classification methods that use high level machine learning techniques. By using these biomarkers, classification technics can be used to predict if a subject has a specific PD syndrome. In short, classification technics utilize a set of "input features", which are essentially syndrome specific biomarkers. Most classification methods use "training" and "testing" datasets with relevant input features. The training datasets are predetermined true cases where the user has explicitly declared which datasets corresponds to which syndrome(s). The classification algorithm will use these true instances to predict the syndromes associated with new datasets which are unknown to the classifier. Important metrics which are used to evaluate the classifiers validity include accuracy, sensitivity, specificity and others. Common classification methods include Support Vector Machines (SVM), k-Nearest Neighbors (K-NN), Binary Logistic Regression, Random Forest (RF), Linear Discriminant Analysis and Bayesian Networks (BN). A short description of several classifiers is given below.

In Support Vector Machines (SVM), an N-dimensional hyperplane is used to optimally separate the dataset into two data classes[31]. SVM deals with a margin on either side of the hyperplane that separates the datasets. Maximizing this margin creates the largest possible distance between the hyperplane and the instances which in turn reduces the generalization error. The model complexity of SVM is not affected by the number of input features in the training set, meaning that SVM is used best for datasets with large number of features. However, many classification problems include non separable data. A common solution for this problem is to map the data onto a higher dimensional space and define a hyperplane there. This new transformed space is called the transformed feature space. After this step, a "kernel function" is used to map the new points into the feature space. The kernel function should be chosen with attention since it directly deals with the transformed feature space in which the training instances will be classified.

The k-nearest neighbor (K-NN) algorithm classifies objects on closest training examples in the feature space. An object is classified by a majority vote of its neighbors, with the object being assigned to the class is most common amongst its "k" nearest neighbors. These neighbors are chosen from a group of objects where the correct classification is known. In fact, K-NN locates the k nearest instances to the unknown instance and determines its class by identifying the single most frequent class label. In a more broader sense, instances can be considered as points within an n-dimensional instance space where each of the n-dimensions corresponds to one of the n-features that are used to describe an instance. Several limitations of this method include, large storage requirements and computational expensiveness, sensitivity to the similarity function (The function that is used to compare instances) and permutational "k" selection[31].

In Binary Logistic Regression, the class label or the target is considered categorical. For example, in PD datasets, the categories may consist of whether the subject is affected by IPD or not. In fact, logistic regression can be used only with two types of target variables. A categorical target variable that has exactly two categories (binary or dichotomous variable) and a continuous target variable that has values

in the range of 0.0 to 1.0 representing probability values or proportions[31].

Random Forest (RF) is an ensemble classifier meaning that it consists of several decision trees. The output of RF is the class which is the mode of the class's output in each individual tree. In detail, decision trees are trees that classify instances by sorting them based on feature values. Each node in a decision tree represents a feature in an instance to be classified, and each branch represents a value that the node can assume. Random Forest classifiers have several important benefits such as, easy of use and implementation, high levels of accuracy, independent testing phase without the need for additional datasets and high resistance to overfitting[31].

Linear Discriminant Analysis, tries to find a linear transformation of two specific predictors which results in a new set of transformed values. LDA and the similar Fisher's Linear Discriminant are easy to implement methods that find the best linear combination of features which separate two or more classes optimally. LDA is more commonly used with continuous measurements whereas Discriminant Correspondence Analysis (DCA) is used in categorical instances[31].

Bayesian Networks (BN), deal with the probability relationships among sets of variables. In this network structure, each node corresponds to a specific feature. BN has several limitations namely, high computational cost meaning that they are not well suited for datasets with high features[31].

In this research project, many of the aforementioned classifiers and other classification methods will be tested and their discriminatory abilities evaluated. In all the classifiers that are used, the input features will be syndrome specific biomarkers. For example, in IPD there is slight change in cortical thickness and volume compared to healthy controls. This specific feature, among others, can be used in a classifier to differentiate IPD and HC. In the first step, this research project aims to find relevant biomarkers which are specific to sub syndromes of PD. These markers will be extracted from multi model MRI dataset of patients with different PD syndromes and healthy controls. After evaluating these markers and selecting important features, several classification technics will be tested using these input features and their accuracies will be evaluated. The final aim of this project is to build an accurate

and robust software that could potentially assists clinicians in their diagnosis of PD patients and also decrease the high rate of miss diagnosis for this disease. More information on how this system operates will be discussed in the subsequent sections.

II. Hypothesis and Aims

Several studies in the past have tried to use syndrome specific biomarkers extracted from single channel Magnetic Resonance Imaging (MRI) images to classify PD syndrome and healthy controls. For instance, (Péran et al., 2010)[32] presented a multi parametric logistic regression method to classify IPD and HC using image-based features of iron deposit, atrophy, and micro structural damage. Therefore, the Thalamus, Putamen, Caudate, Pallidum, Substantia Nigra (SN), and red nucleus were semi-automatically segmented in the high-resolution T1-weighted MRI dataset employing the FSL software and used to analyze the T2*-weighted and Diffusion Tensor Imaging (DTI) MRI datasets. More precisely, the DTI datasets were used to calculate parametric maps of fractional anisotropy (FA) and mean diffusivity (MD), while the multi-echo T2*-weighted datasets were used to calculate quantitative R2* maps. All three parametric maps were then registered to the T1-weighted dataset to calculate average values for each parameter and brain region. Additionally, the volume of each structure was also calculated. Training and testing of the logistic regression was conducted using a database of 30 IPD, and 22 HC subjects and a 10 fold cross validation. It was found that the combination of mean R2* values in the left or right SN, FA in the right SN and MD in the putamen or caudate nucleus leads to the best discriminating power resulting in a global accuracy of more than 95%. Similar studies like the one mentioned have been proposed which make use of single or in the best case scenario double channel MRI datasets. These single channel approaches will generally result in less favorable accuracies which are not accurate enough for real clinical applications. In this research project, we hypothesize that datasets acquired form multi channel MRI can potentially result in an increased classification accuracy for different PD syndromes and healthy controls.

The broad aims of this research project can be summarized in the following three goals:

Aim 1: Identification of relevant and discriminative biomarkers using multi channel MRI datasets

Aim 2: Testing of different classifiers by using the input features obtained from Aim 1 and evaluating their accuracies

Aim 3: Developing a diagnosis aid classification software with a fully automatic pipeline

III. Materials & Methods

Several classification technics (such as SVM, Random Forest, etc) will be implemented and tested using the multi spectral data obtained from multi channel MRI datasets. The datasets in this research project includes, 76 IPD patients, 22 PSP patients, and 40 age-matched healthy control subjects acquired at the University Medical Center Hamburg-Eppendorf, Germany, using a 3T Siemens Skyra MR scanner. As it was mentioned before, unlike previous studies, the novelty of this project is based on the usage of multi spectral information for Parkinson's disease classification. In this section, an overview of some modalities that have used classifiers for PD syndrome differentiation are given.

III.A) T1-weighted imaging

T1 weighted volumetric MRI sequences, are commonly used in clinical MRI routines[33]. One of the most important features of T1 weighted imaging is the high grey/white matter contrast which is useful for segmentation of anatomical or functional subregions of the brain. Using this imaging modality, previous studies have shown that the cortex undergoes a slight reduction in volume and thickness in patients diagnosed with IPD[5]. However, the contrast of T1 imaging is often not suitable to identify important PD related brain structures such as SN. This is supported by previously contradicting studies that evaluated atrophy measurements by using voxel based morphometry (VBM) in IPD versus healthy

controls (HC) by looking at the caudate nucleus[34,35] and hippocampus[35]. Several studies in the past have used biomarkers obtained through T1-weighted imaging and have implemented them in the classification of PD subtypes versus controls. (Federico Nemmi et al. 2015)[33] used (1) gray matter density and (2) subcortical nuclei volume and (3) brain region shapes to classify 21 IPD and 20 healthy controls (HC). They performed linear discriminant analysis on the two groups using both global volume and shape only differences (i.e. Gray matter density, subcortical nuclei volume and subcortical nuclei shape). Moreover, DTI was used to analyze structural connectivity distribution between local atrophy and cortical regions. Brain segmentation was conducted using FIRST[36] in subcortical nuclei such as amygdala, caudate nucleus, hippocampus, globus pallidus, putamen, left and right nucleus accumbens and thalamus.. Linear discriminant analysis with leave one out validation method was used for both (1) several brain region shapes and (2) global volume combinations such as left putamen, left caudate or left caudate and right putamen to evaluate classification accuracy. For the volume only analysis, the classifier reached accuracies up to 65% for both the left putamen and the combination of left caudate and left putamen. In the case of shape only analysis, the reported accuracies were up to 83% for the left putamen and also the combination of left caudate and left putamen. The study concluded that subcortical shape results (local atrophy-only) were able to better categorize PD and controls compared to standard volumetric-only analysis. (Salvatore et al., 2014)[37] utilized morphological T1-weighted images to classify 28 IPD, 28 PSP and 28 HC. They found that voxel based pattern distribution of brain structural differences such as midbrain, pons, corpus callosum and thalamus were most helpful in differentiating PSP and PD. Principal Component Analysis was used as the feature extraction method and SVM was used for the classification of subjects. Their proposed method was validated with (1) leave one out validation (LOO) and (2) two fold cross validation. By using LOO validation the reported overall mean accuracy for PD vs. control, PSP vs. control and PSP vs. PD where 85.8%,89.1% and 88.9%, respectively. By using the two fold cross validation method, the overall mean accuracies were 83.2%, 86.2% and 84.7 for PD vs. control, PSP vs. control and PSP vs. PD, respectively. In a recent

study (Singh et al., 2015)[38], used a combination of least square support vector machine (LS-SVM) and Kohonen self organizing map (KSOM) as a feature extraction method to classify an impressive number of subjects including, 518 IPD, 245 healthy controls (HC) and 68 IPD patients with Scans Without Evidence of Dopaminergic Deficit (SWEDD). In this study, they divided the dataset into Age-Related (ARS) and Age-Unrelated subgroups (AUG) and performed binary classification on combinations of PD, SWEDD and HC based on "tissue by tissue comparison in WM or GM". Automatic image segmentation was conducted by using SPM8. In order to evaluate their classifier, the dataset was randomly divided, where 80% of the data was used for training and 20% was used for testing in each of the binary classification instances. Using AUG, the classifier achieved an average classification accuracy of 95.04%, 99.6% and 99.22% for PD vs. HC, SWEDD vs. HC and PD vs. SWEDD, respectively. The average classification accuracy is obtained by averaging results for GM and WM. Using ARS, in the case of PD vs. HC, SWEDD vs. HC and PD vs. SWEDD, the average accuracy was 99.78%, 100.00% and 100.00%. For classifying PD vs. HC certain GM regions such as medial dorsal nucleus, putamen, pulvinar, Brodmann areas 29 and 23 and for WM, corpus callosum and limbic cortex were helpful. In the case of SWEDD vs. HC, changes in corpus callosum and pulvinar region were observed. Also compared to PD vs. HC, lateral posterior and anterior nucleuses of thalmus were affected. However, in PD vs. SWEDD, the authors couldn't find a significant degeneration of putamen.

III.B) T2-weighted imaging

Histopathological studies have revealed that different Parkinson diseases are associated with varying regional brain iron accumulation in the brain[39]. Non-heme iron such as ferritin and hemosiderin create local magnetic field inhomogeneities producing intra-voxel dephasing and shortened transverse relaxation times, which can be measured using T2-weighted MRI[20]. In previous groups-wise statistics, IPD has been shown to be accompanied with reduced T2 and T2* relaxation times[5]. In MSA-P and

PSP, T2 times have been shown to be affected in brain regions such as striatum and globus pallidus[40]. In PSP, reduced T2 times were observed in basal ganglia, the midbrain, the superior cerebullar peduncle and white matter.[5] Furthermore, (Boelmans et al. 2012)[41]proposed a linear discriminant method to classify IPD, PSP, and HC subjects using quantitative T2' values determined from triple-echo T2 and T2* MRI sequences. Average T2' values were manually measured in multiple deep gray matter and white matter structures. Training and testing of the stepwise linear discriminant analysis classifier using the average T2' values in the caudate, pallidum, putamen, and thalamus as input features was conducted using a database of 30 IPD, 12 PSP, and 24 HC subjects. The results revealed that this method is capable of classifying patients to the three groups with an overall accuracy of 74.2%. Of the 24 HC subjects, 18 were classified correctly whereas 6 subjects were grouped in the IPD group. In the clinically diagnosed IPD group, 9 and 2 patients were wrongly grouped as HC and PSP, respectively. All PSP patients were classified correctly. Moreover, statistical analysis of the T2' values in the analyzed brain regions identified shortened T2' values in caudate nucleus, globus pallidus and putamen in PSP compared to HC and IPD.

III.C) Diffusion Weighted imaging

Diffusion-weighted MR imaging (DWI) and diffusion tensor MR imaging (DTI) are sequences sensitive to water diffusion in the brain tissue. DWI scans can be used to calculate the quantitative apparent diffusion coefficient (ADC), while DTI can be used to calculate the quantitative mean diffusivity (MD), fractional anisotropy (FA), and radial diffusivity (RD), which represent important diffusion parameters that can be used to differentiate PD syndromes. For example, a reduction in FA is commonly present in aging tissue or neurodegenerative illnesses such as PD. Moreover, studies have shown that ADC can differentiate PSP and MSA-P from PD with high levels (90%>) of sensitivity and specificity[42]. (Salamanca et al. 2015)[43]proposed a method to classify IPD and HC subjects using a combination of a

fisher vectors and logistic linear regression based on FA and MD values in 14 regions of interest as features. Therefore, the FA and MD parameter maps were generated from the DTI data. After this, the MNI brain atlas was non-linearly registered to the DTI data and the resulting non-linear transformation field was used to warp 14 regions of interest from the MNI atlas to the patient space to extract the corresponding FA and MD values. Using a database of 100 patients (50 IPD and 50 HC), the proposed method was evaluated using 100 experiments of 10-fold cross-validation resulting in a maximal accuracy of 77%. (Haller et al. 2012)[44] presented an approach to classify IPD subjects and patients with atypical forms of Parkinsonism (APS) using a SVM as the classifier and voxel-wise FA values as features. Briefly, FA parameter maps were generated from DTI data and non-linearly registered to mean FA brain atlas. After this, each voxel of the brain tissue was used as a feature for the classification using an SVM with RBF kernel based on the FA values. To reduce the number of features for the classification, a feature selection was performed using the RELIEFF[45] algorithm separately with the most discriminative 100, 250, 500, 750 and 1000 features (voxels) used for SVM classification. Each SVM classification model with varying number of training features was evaluated using a 10-fold cross validation. The best classification result with an accuracy of 97.5% was achieved for the SVM using the top 100 features. Voxel-based statistics revealed a significant increase in FA and a decrease in MD and RD in the right frontal WM in IPD compared to HC.

I.V. Analysis & Statistics

There are several metrics that evaluate the performance of classifiers. Some of the most important evaluation metrics are sensitivity, specificity, accuracy, kappa statistics (KS), precision, f- measure and area under the receiver operating characteristics (AUC). Many classification problems in medical applications deal with binary (two case) instances. For example, in this project a subject either belongs to the IPD group or healthy control (By extension, PSP vs. IPD or IPD vs. HC). The overall objective of

the classifier is to place each subject in the corresponding correct group. Here the "correct" group is defined based on the real clinical diagnosis of each subject. The possible outcomes of such classification results are defined in statistical terms such as true positive (TP), false positive (FP), true negative (TN) and false negative (FN). Sensitivity or true positive rate (TPR) is defined as TP divided by the summation of TP and FN whereas, specificity or true negative rate (TNR) is defined as TN divided by the summation of TN and FP. In detail, sensitivity depicts the number of subjects with a specific PD syndrome who have been correctly identified as such by the classifier. Specificity, denotes the percentage of healthy controls who have been correctly grouped as non-syndrome subject by the classifier. Accuracy is defined by the summation of TP and TN divided by all the possible outcomes. ROC curves are plotted using TPR versus false positive rate, as the discrimination threshold of the classification method is changed. Furthermore, Kappa error which is used to compare the performances of classifiers, is a well suited method to analyze classifications based on chance. F-measure is defined as the harmonic mean of sensitivity and precision whereas, precision is defined as TP divided by the summation of TP and FP. In this research project, the aforementioned metrics will be used to evaluate classifier performance.

V. Project Overview

In this section, a broad overview of the project and all of its main components will be presented. Figure 5.1 depicts the main sections of this project.

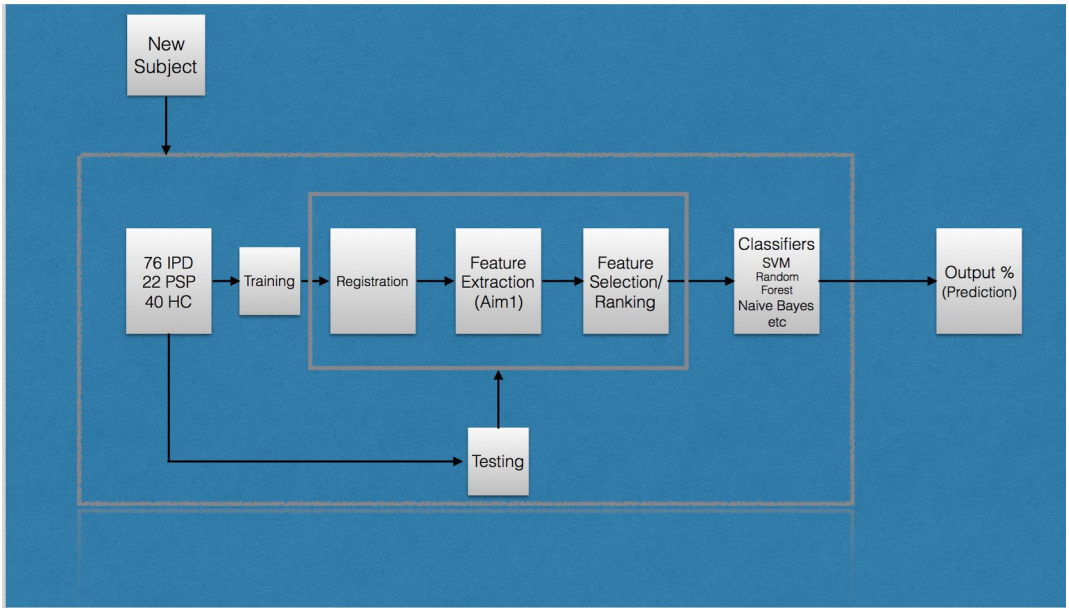

Figure 5.1. A schematic representation of this project

In the first step, the dataset will be split into two groups of training and testing. As it was mentioned before, the training set will be used to train the classifier while the testing dataset will validate it. After this step, the images in the training set will be registered using a specific atlas. Registration is defined as the determination of a geometrical transformation that aligns points in one view of an object with corresponding points in another view of that object or another object. In detail, image registration is the process of aligning two or more images of the same scene where one of the images is designated as the reference images and a geometrical transformation is applied on the other image to align it with the reference image. There are two types of transformation, namely rigid and non-rigid. In this project, an affine transform which is a non-rigid transformation will be used to register the datasets to the Montreal Neurological Institute (MNI) atlas. In the next step, important discriminative features will be extracted from the registered images. The information regarding this phase of the project is obtained form the

first overall aim of the project. For example, the study by (Tyler Rolheiser et al., 2011)[13] suggests that fractional anisotropy (FA) in the olfactory region is a distinguishing factor between IPD and healthy controls. (Gupta et al., 2010)[24] showed that hypo intensity of red nucleus has the ability to differentiate PSP and MSA-P. Moreover, (Ghaemi et al., 2002)[46] proposed that decreased putaminal volume was a sign that could discriminate between MSA-P and IPD. Furthermore, tissue reduction in cerebral peduncles and midbrain has been shown by (Price et al., 2004)[47] to be a differentiating sign in PSP vs. IPD as well as PSP vs. healthy controls. Similar research have found more syndrome specific signs which can be considered as relevant biomarkers meaning that these unique features can be used as input features in classifiers. The next step, is the feature ranking phase of the project. Generally, steps such as feature selection and feature ranking are considered essential aspects in most classification technics. Feature ranking, as the name suggests, deals with ranking the obtained features from the previous step in order of their discriminatory power. In fact, adding irrelevant features to the classifier will almost always decrease the classification accuracy. There are several feature ranking approaches, namely embedded technics where feature ranking is an internal part of the classification method. The second approach is the filtered method where the features are pre ranked before the classification is performed. For example, Fisher filtering is used to rank input features based on their overall importance. The third method, which is called the wrapper approach, a classification technic is used to identify the "best" subset of features. In general, data are better characterized with fewer variables, therefor, the most important goal here is to minimize the error rate with the minimum features of PD datasets. After the feature selection/ranking phase, the input features are ready to be used as the training data for the classifiers. As it was mentioned before, we will use a wide range of classification algorithms and evaluate their performance. The evaluation will be conducted by the testing dataset. It is important to consider that classification methods are inherently permutational, meaning that we need to "tune" the parameters and observe changes in the classification accuracy (The most important evaluation metrics are referred to in section 4). However, background knowledge on the different types

of classification methods and their overall advantages and disadvantages in different classification problems, helps us to choose the "right" classifiers. After all the aforementioned steps are implemented, the second aim of the project will be fulfilled. As for the third step, the proposed system will be wrapped as a single software package. Since one of the overall goals of this project is to assists clinicians in PD diagnosis, the software needs to be tested by using completely new and unknown datasets. These datasets consist of multispectral MRI images from patients with unknown PD syndromes. When the software is confronted with a new dataset, it will produce a syndrome prediction. This prediction combined with the clinicians own initial diagnosis can help reduce the rate of PD miss diagnosis.

1. Mhyre, T. R., Boyd, J. T., Hamill, R. W. & Maguire-Zeiss, K. A. *Protein Aggregation and Fibrillogenesis in Cerebral and Systemic Amyloid Disease*. *Sub-cellular biochemistry* **65,** (2012).

2. FEARNLEY, J. M. & LEES, A. J. Ageing and Parkinson'S Disease: Substantia Nigra Regional Selectivity. *Brain* **114,** 2283–2301 (1991).

3. Gattellaro, G. *et al.* White matter involvement in idiopathic Parkinson disease: A diffusion tensor imaging study. *Am. J. Neuroradiol.* **30,** 1222–1226 (2009).

4. Tambasco, N. *et al.* Magnetization transfer changes of grey and white matter in Parkinson's disease. *Neuroradiology* **45,** 224–30 (2003).

5. Pyatigorskaya, N., Gallea, C., Garcia-Lorenzo, D., Vidailhet, M. & Lehericy, S. A review of the use of magnetic resonance imaging in Parkinson's disease. *Ther. Adv. Neurol. Disord.* **7,** 206–220 (2013).

6. Morelli, M. *et al.* Accuracy of magnetic resonance parkinsonism index for differentiation of progressive supranuclear palsy from probable or possible Parkinson disease. *Mov. Disord.* **26,** 527–533 (2011).

7. Zhu, M.-W., Liu, J., Arzberger, T., Wang, L.-N. & Wang, Z.-F. Typical or atypical progressive supranuclear palsy: a comparative clinicopathologic study of three Chinese cases. *Int J Clin Exp Pathol* **8,** 867–874 (2015).

8. Wenning, G., Colosimo, C., Geser, F. & Poewe, W. Multiple System Atrophy. *Lancet Neurol.* **3,** 93–103 (2003).

9. Lipp, A. *et al.* Prospective differentiation of multiple system atrophy from Parkinson disease, with and without autonomic failure. *Arch. Neurol.* **66,** 742–50 (2009).

10. Kume, A., Takahashi, A. & Hashizume, Y. Neuronal cell loss of the striatonigral system in multiple system atrophy. *J. Neurol. Sci.* **117,** 33–40 (1993).

11. Erbetta, A. *et al.* Diffusion tensor imaging shows different topographic involvement of the thalamus in progressive supranuclear palsy and corticobasal degeneration. *Am. J. Neuroradiol.* **30,** 1482–1487 (2009).

12. Yekhlef, F. *et al.* Routine MRI for the differential diagnosis of Parkinson's disease, MSA, PSP, and CBD. *J. Neural Transm.* **110,** 151–169 (2003).

13. Rolheiser, T. M. *et al.* Diffusion tensor imaging and olfactory identification testing in early-stage Parkinson's disease. *J. Neurol.* **258,** 1254–1260 (2011).

14. Statistics on Parkinson's - Parkinson's Disease Foundation (PDF). Available at: http://www.pdf.org/en/parkinson_statistics. (Accessed: 26th March 2016)

15. Parkinson's disease: Prevalence, diagnosis and impact. (2014).

16. Hughes, A. J., Daniel, S. E., Kilford, L. & Lees, A. J. Accuracy of clinical diagnosis of idiopathic

Parkinson ' s disease : a clinico-pathological study of 100 cases. 181–184 (1992). doi:10.1136/jnnp.55.3.181

17. Ene, M. Neural network-based approach to discriminate healthy people from those with Parkinson's disease. *Ann. Univ. Craiova, Math. Comp. Sci. Ser.* **35,** 112{116 (2008).

18. Singh, N., Pillay, V. & Choonara, Y. E. Advances in the treatment of Parkinson's disease. *Prog. Neurobiol.* **81,** 29–44 (2007).

19. Tang, C. C. *et al.* Differential diagnosis of parkinsonism: a metabolic imaging study using pattern analysis. *Lancet Neurol.* **9,** 149–158 (2010).

20. Forkert, N. D. *et al.* Automated volumes-of-interest identification for classical and atypical parkinsonian syndrome differentiation using T2' MR imaging. *Methods Inf. Med.* **52,** 128–136 (2013).

21. Duchesne, S., Rolland, Y. & V??rin, M. Automated Computer Differential Classification in Parkinsonian Syndromes via Pattern Analysis on MRI. *Acad. Radiol.* **16,** 61–70 (2009).

22. Messina, D. *et al.* Patterns of brain atrophy in Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. *Parkinsonism Relat. Disord.* **17,** 172–176 (2011).

23. Forkert, N. D. *et al.* Image-based classification of parkinsonian syndromes using T2'-atlases. *Stud. Health Technol. Inform.* **169,** 465–469 (2011).

24. Gupta, D., Saini, J., Kesavadas, C., Sarma, P. S. & Kishore, A. Utility of susceptibility-weighted MRI in differentiating Parkinson's disease and atypical parkinsonism. *Neuroradiology* **52,** 1087–1094 (2010).

25. Tir, M., Delmaire, C., Besson, P. & Defebvre, L. The value of novel MRI techniques in Parkinson-plus syndromes: diffusion tensor imaging and anatomical connectivity studies. *Rev. Neurol. (Paris).* **170,** 266–76 (2014).

26. Egerton, T., Williams, D. R. & Iansek, R. Comparison of gait in progressive supranuclear palsy, Parkinson's disease and healthy older adults. *BMC Neurol.* **12,** 116 (2012).

27. Drot??r, P. *et al.* Analysis of in-air movement in handwriting: A novel marker for Parkinson's disease. *Comput. Methods Programs Biomed.* **117,** 405–411 (2014).

28. Kugler, P. *et al.* Automatic recognition of Parkinson's disease using surface electromyography during standardized gait tests. *Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.* **2013,** 5781–4 (2013).

29. Casjens, S. *et al.* Diagnostic Value of the Impairment of Olfaction in Parkinson's Disease. *PLoS One* **8,** 1–7 (2013).

30. DeMarshall, C. A. *et al.* Potential utility of autoantibodies as blood-based biomarkers for early detection and diagnosis of Parkinson's disease. *Immunol. Lett.* **168,** 80–88 (2015).

31. Ramani, R. G. Parkinson Disease Classification using Data Mining Algorithms. **32,** 17–22 (2011).

32. Péran, P. *et al.* Magnetic resonance imaging markers of Parkinson's disease nigrostriatal signature. *Brain* **133,** 3423–3433 (2010).

33. Nemmi, F., Sabatini, U., Rascol, O. & Péran, P. Parkinson's disease and local atrophy in subcortical nuclei: Insight from shape analysis. *Neurobiol. Aging* **36,** 424–433 (2015).

34. Brenneis, C. *et al.* Voxel-based morphometry detects cortical atrophy in the Parkinson variant of multiple system atrophy. *Mov. Disord.* **18,** 1132–1138 (2003).

35. Tessitore, a *et al.* Regional gray matter atrophy in patients with Parkinson disease and freezing of gait. *AJNR. Am. J. Neuroradiol.* **33,** 1804–9 (2012).

36. Patenaude, B., Smith, S. M., Kennedy, D. N. & Jenkinson, M. A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage* **56,** 907–922 (2011).

37. Salvatore, C. *et al.* Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and Progressive Supranuclear Palsy. *J. Neurosci. Methods* **222,** 230–237 (2014).

38. Singh, G. & Samavedham, L. Unsupervised learning based feature extraction for differential diagnosis of neurodegenerative diseases: A case study on early-stage diagnosis of Parkinson disease. *J. Neurosci. Methods* **256,** 30–40 (2015).

39. Schenck, J. F. & Zimmerman, E. A. High-field magnetic resonance imaging of brain iron: Birth of a biomarker? *NMR Biomed.* **17,** 433–445 (2004).

40. Wang, Y. *et al.* Different iron-deposition patterns of multiple system atrophy with predominant parkinsonism and idiopathetic Parkinson diseases demonstrated by phase-corrected susceptibility-weighted imaging. *Am. J. Neuroradiol.* **33,** 266–273 (2012).

41. Boelmans, K. *et al.* Brain iron deposition fingerprints in Parkinson's disease and progressive supranuclear palsy. *Mov. Disord.* **27,** 421–427 (2012).

42. Schocke, M. F. H. *et al.* Diffusion-weighted MRI differentiates the Parkinson variant of multiple system atrophy from PD. *Neurology* **58,** 575–80 (2002).

43. Xia, W. *et al.* Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. *MICCAI 2015, Part I, Lect. Notes Comput. Sci.* **9349,** 371–378 (2015).

44. Haller, S. *et al.* Individual detection of patients with Parkinson disease using support vector machine analysis of diffusion tensor imaging data: initial results. *AJNR. Am. J. Neuroradiol.* **33,** 2123–8 (2012).

45. Kononenko, I., Šimec, E. & Robnik-Šikonja, M. Overcoming the myopia of inductive learning algorithms with RELIEFF. *Appl. Intell.* **7,** 39–55 (1997).

46. Ghaemi, M., Hilker, R., Rudolf, J., Sobesky, J. & Heiss, W.-D. Differentiating multiple system

atrophy from Parkinson's disease: contribution of striatal and midbrain MRI volumetry and multi-tracer PET imaging. *J. Neurol. Neurosurg. Psychiatry* **73,** 517–23 (2002).

47. Price, S. *et al.* Voxel-based morphometry detects patterns of atrophy that help differentiate progressive supranuclear palsy and Parkinson's disease. *Neuroimage* **23,** 663–669 (2004).