International Journal of Medical Informatics 90 (2016) 13–21

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.ijmijournal.com

High-Accuracy Detection of Early Parkinson's Disease through Multimodal Features and Machine Learning

R. Prashanth [a,*], Sumantra Dutta Roy [a], Pravat K. Mandal [b,c], Shantanu Ghosh [d]

[a] Department of Electrical Engineering, Indian Institute of Technology, Delhi, India
[b] Neuroimaging and Neurospectroscopy Laboratory, National Brain Research Centre, India
[c] Department of Radiology, Johns Hopkins Medicine, MD, USA
[d] Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, MA, USA

ARTICLE INFO

Article history:
Received 3 December 2015
Received in revised form 4 February 2016
Accepted 1 March 2016

Keywords:
Parkinson's disease
non-motor features
cerebrospinal fluid markers
dopaminergic imaging
Computer-aided diagnosis
pattern classification

ABSTRACT

Early (or preclinical) diagnosis of Parkinson's disease (PD) is crucial for its early management as by the time manifestation of clinical symptoms occur, more than 60% of the dopaminergic neurons have already been lost. It is now established that there exists a premotor stage, before the start of these classic motor symptoms, characterized by a constellation of clinical features, mostly non-motor in nature such as Rapid Eye Movement (REM) sleep Behaviour Disorder (RBD) and olfactory loss. In this paper, we use the non-motor features of RBD and olfactory loss, along with other significant biomarkers such as Cerebrospinal fluid (CSF) measurements and dopaminergic imaging markers from 183 healthy normal and 401 early PD subjects, as obtained from the Parkinson's Progression Markers Initiative (PPMI) database, to classify early PD subjects from normal using Naïve Bayes, Support Vector Machine (SVM), Boosted Trees and Random Forests classifiers. We observe that SVM classifier gave the best performance (96.40% accuracy, 97.03% sensitivity, 95.01% specificity, and 98.88% area under ROC). We infer from the study that a combination of non-motor, CSF and imaging markers may aid in the preclinical diagnosis of PD.

1. Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder immensely affecting the quality of lives of millions of people worldwide [1]. It is caused by the loss of dopaminergic neurons in the substantia nigra and due to which there is substantial decrease of dopamine content in the striatum, and the amount of dopamine transporters. The clinical diagnosis of PD is based on the presence of four cardinal motor symptoms which are tremor at rest, rigidity, bradykinesia and postural instability. These symptoms start appearing only when there is substantial reduction of about 60% of dopaminergic neurons (or about 80% reduction in the dopamine concentration in the putamen) [1,2]. As of now, there is only symptomatic treatment available for PD and initiation of treatment at a later stage is of little help as the deterioration becomes extensive. Hence, early detection of PD is crucial for early management and for allowing neuroprotective strategies, to be administered earlier in the disease process, when available.

It is now established that there exists a time-span called the premotor or prodromal phase in PD, which lasts at least for 5 years and possibly for 20 years, between the onset of neurodegeneration and manifestation of classic clinical motor symptoms [3]. During this phase, the subject mostly shows non-motor symptoms such as Rapid Eye Movement (REM) sleep Behavior Disorder (RBD) and olfactory loss. However, none of these symptoms carry sufficient sensitivity that can be used for screening. In this case, they may be used in conjunction with other potential biomarkers such as Cerebrospinal fluid (CSF) measurements and dopamine transporter imaging for identifying subjects (or population) at risk of PD [3–5]. A brief description of these features is given below:

1.1. Non-motor features

Non-motor features affect the quality of life equally or more than the motor features in PD. Among these non-motor features, hyposmia (olfactory dysfunction) and RBD are the most common ones.

Abbreviations: PD, Parkinson's disease; PPMI, Parkinson's Progression Markers Initiative; REM, Rapid Eye Movement; RBD, REM sleep Behaviour Disorder; RBDSQ, RBD Screening Questionnaire; UPSIT, University of Pennsylvania Smell Identification Test; CSF, Cerebrospinal Fluid; Aβ1-42, amyloid beta peptide 1-42; α-Syn, α-synuclein; P-tau$_{181}$, tau phosphorylated at threonine 181; T-tau, total tau; SPECT, Single Photon Emission Computed Tomography; SVM, Support Vector Machine; SBR, Striatal Binding Ratio; SBR$_{RC}$, Striatal Binding Ratio for right caudate; SBR$_{LC}$, Striatal Binding Ratio for left caudate; SBR$_{RP}$, Striatal Binding Ratio for right putamen; SBR$_{LP}$, Striatal Binding Ratio for left putamen.

* Corresponding author. Present address: R & D Engineer, Azoi Mobile Technologies Pvt. Ltd, Ahmedabad 380009, India.
E-mail address: prashanth.r.itd@gmail.com (R. Prashanth).

http://dx.doi.org/10.1016/j.ijmedinf.2016.03.001
1386-5056/

Studies have shown that profound olfactory deficits in the form of impairments in odor detection, differentiation, and identification are observed at the earliest clinical stage of PD indicating that olfactory dysfunction is a prodromal or preclinical sign [6]. Ponsen et al. performed a study on a cohort of 361 asymptomatic relatives of PD patients, and observed that idiopathic olfactory dysfunction is associated with an increased risk of developing PD of at least 10% [6].

RBD is another non-motor symptom which often antedates neurodegenerative disorders such as PD [7]. It is a disturbance associated with sleep and is characterized by vivid, aggressive or action-packed dreams, dream enacting behaviours such as shouting, punching, and loss of normal REM-sleep muscle atonia [7]. Studies show that subjects diagnosed with RBD are at high risk of developing neurodegenerative diseases, most importantly PD [7,8]. Postuma et al. performed a clinical follow-up study for 12 years on 93 subjects with RBD, and observed that over this period, 26 (28%) developed neurodegenerative disorders of which 14 (15%) were PD. From this study, they also estimated the 5-year risk, 10-year risk and 12-year risk of neurodegenerative disease in idiopathic RBD as 17.7%, 40.6% and 52.4% respectively [8].

1.2. Cerebrospinal fluid (CSF) markers

CSF markers of amyloid beta peptide 1–42 (Aβ1-42), total tau (T-tau), tau phosphorylated at threonine 181 (P-tau$_{181}$) and α-synuclein (α-Syn) are now widely researched biomarkers for the diagnosis of neurodegenerative disorders [9]. CSF have the advantage that they are more accessible, less costly than imaging, reflect metabolic and pathological states of the central nervous system (CNS) more directly than any other body fluids [10]. Recent studies have shown that CSF markers hold promise in detecting PD, but they are still in their early stages of development [2]. For instance, Kang et al. performed a study with 63 early PD and 39 healthy normal subjects from the Parkinson's Progression Markers Initiative (PPMI) database, and observed a significant decrease in the levels of Aβ1-42, T-tau, P-tau$_{181}$, α-Syn and T-tau/Aβ1-42 in PD as compared to healthy normal indicating their prognostic and diagnostic potential in early-stage PD [11]. However, they also observed that the diagnostic utility of these CSF markers is low as the area under the ROC curve (AUC) were less than 80% and that a combination with other significant biomarkers is likely to improve the diagnostic accuracy.

1.3. Neuroimaging markers

Dopamine transporter imaging using Single Photon Emission Computed Tomography (SPECT) is becoming a routine use at least in cases where there is diagnostic uncertainty with regard to neurodegenerative parkinsonism [2]. An abnormal imaging strongly supports a diagnosis of PD or other neurodegenerative striatonigral disease, while a normal scan for a patient with symptoms is suggestive of a diagnosis other than neurodegenerative parkinsonism [2]. Studies have demonstrated that SPECT has potential to depict dopaminergic transporter (DAT) loss, even in the subclinical stages of PD [2,12]. For instance, SPECT imaging have shown dopamine transporter reduction in subjects without PD but with hyposmia and RBD which are considered very common premotor or preclinical symptoms in PD. Berendse and Ponsen [13] in their study observed that 4 out of 25 hyposmic relatives of PD showed subclinical reductions in dopamine transporter binding as depicted in their SPECT scans, and two of whom subsequently developed clinical parkinsonism, and none of the 23 normosmic (normal olfactory function) relatives of PD showed abnormal SPECT scans. In another study by Stiasny-Kolster et al., they observed that 3 out of 11 patients with 'idiopathic' clinical RBD showed reduced striatal DAT binding, and one of these had clinical parkinsonism [4].

There are studies which use these features to classify subjects as early PD and healthy normal [11,14–22]. [14] performed smell identification tests using culturally adapted translations of the University of Pennsylvania Smell Identification Test (UPSIT) and Sniffin' Sticks (SS) test in 106 PD and 118 normal subjects. They observed the following performance measures using logistic regression for classification. Using SS test: 85.3% accuracy, 81.1% sensitivity and 89% specificity; and using UPSIT: 82.8% accuracy, 82.1% sensitivity and 83.5% specificity. [15] improved the same study by including more subjects, from 193 PD and 157 normal subjects, and using logistic regression, they observed an enhanced classification accuracy with both SS test and UPSIT with highest performance obtained with the SS test as 88.4% accuracy, 90.4% sensitivity and 85.5% specificity. [16] used non-motor features such as cognitive impairment, psychiatric complications, autonomic dysfunction or sleep disturbance, from 410 PD patients, to classify subjects based on disease severity as mild, moderate and severe, and they obtained classification accuracies in the range of 72%–92%. [11] used Cerebrospinal Fluid (CSF) measurements from 63 early PD and 39 healthy normal and observed that these measures were statistically significant but showed a low diagnostic utility as the area under the ROC (AUC) was less than 0.8 for PD diagnosis. [17] used 123I-Ioflupane (DaTSCAN^{TM}, GE Healthcare; also known as [^{123}I]FP-CIT) SPECT scan data from 79 patients with parkinsonism (PS) and 37 non-PS subjects (which is not the healthy normal group and instead it represent subjects with non-PS conditions like essential tremor that shows normal SPECT scans), and obtained a classification accuracy of 94.8% with 93.7% sensitivity and 97.3% specificity using Naïve Bayes classifier. [18] also used SPECT scan data from 95 PD and 94 normal subjects and obtained a maximum accuracy of 94.7%, sensitivity 93.7% and specificity 95.7% using an approach based on partial least squares and support vector machine (SVM). [19] used SPECT scan data from 108 PS and 100 normal subjects and using SVM classifier obtained a sensitivity of 89.02%, specificity as 93.21% and area under the ROC curve (AUC) as 96.81% for classification. [20] used Striatal Binding Ratio (SBR) values from SPECT imaging corresponding to 369 early PD and 179 normal subjects from the PPMI database and obtained a maximum classification accuracy of 96.14% using SVM. [21] carried out speech analysis to extract features for classification between speech symptom severity levels using SVM. They obtained accuracies between 85%–92% with average AUC around 91%. [22] carry out detection of Parkinson's disease (PD) from speech signal measurements using machine learning techniques such as logistic regression, SVM, ensemble methods etc. The data used for the study consisted of 31 normal and 23 PD subjects. They obtained encouraging results. However, these studies had the limitation that none of them used a combination of features for performing the classification or they used smaller sample size.

In this paper, we use a combination of non-motor features of RBD and olfactory loss, CSF measurements and SPECT imaging markers, obtained from the PPMI [9] database, to develop diagnostic models to classify subjects into early PD and healthy normal, using Naïve Bayes, SVM, Boosted trees and Random forests classifiers, which may aid in the early diagnosis of PD.

2. Materials and Methods

2.1. Database

The data used in the study were from the Parkinson's Progression Markers Initiative (PPMI) database (http://www.ppmi-info.org/data). For up-to-date information, please visit http://www.

ppmi-info.org. The PPMI [9] is a landmark, large-scale, comprehensive, observational, international, multi-center study that recruits *de novo* (early-untreated) PD patients and age-matched healthy normal subjects to identify PD progression biomarkers.

We use the University of Pennsylvania Smell Identification Test (UPSIT) data which evaluates olfactory dysfunction, RBD screening questionnaire (RBDSQ) which reflects RBD characteristics, CSF markers of Aβ1-42, α- syn, P-tau$_{181}$, T-tau, T-tau/Aβ1-42, P-tau$_{181}$/Aβ1-42 and P-tau$_{181}$/T-tau, and SPECT measurements of striatal binding ratio (SBR) data from the PPMI database for our study. A brief description of these features is given in Section 2.3.

2.2. Cohort details

PPMI is a longitudinal study where subjects undergo a comprehensive longitudinal follow-up schedule of clinical, imaging and biospecimen (CSF) assessments. The number of normal subjects is 183 and the number of early PD patients is 401, in this study. All the PD subjects are in their early phase i.e., in stage 1 or 2 in the Hoehn and Yahr scale [23] with mean as 1.56 ± 0.50.

2.3. Feature description

2.3.1. University of Pennsylvania Smell Identification Test (UPSIT)

The 40-item UPSIT [24] is one of the most reliable (test-retest $r = 0.94$), accurate and widely used smell identification tests available. It consists of four booklets and each booklet contains 10 pages with each page containing a different odor in a plastic microcapsule. The subject scrapes the strip with a pencil, which releases the odor, and then identifies the smell and marks an option from the four choices in the page that best describes the odor. The maximum score from the test can be 40 when all the odors are correctly identified.

2.3.2. REM sleep Behavior Disorder Screening Questionnaire (RBDSQ)

RBDSQ is a specific screening questionnaire for RBD, developed by Stiasny-Kolster et al. [25], assessing the most prominent clinical features of RBD. Researchers have studied the utility of RBDSQ and observed that it performed with high sensitivity and reasonable specificity [25,26]. RBDSQ consists of a set of questions which is answered with a 'yes ' (1) or 'no' (0) response. We use the total score from first nine items which can be a maximum of 12. Higher scores indicate higher chances of having the disorder.

2.3.3. Biomarkers from Cerebrospinal fluid (CSF)

CSF samples at PPMI were collected as per the PPMI Research Biomarkers Laboratory Manual (Biologics Manual, http://www.ppmi-info.org/study-design/research-documents-and-sops/) and the measurements of Aβ1−42, T-tau and P-tau$_{181}$ were measured using the research-use-only multiplex xMAP Luminex platform and Innogenetics immunoassay kits (Innogenetics/Fujirebio, Ghent, Belgium), as described in [27]. The concentration of α-Syn in CSF samples collected for PPMI were measured with an ELISA assay available commercially from Covance (cat # SIG-38974-kit).

In a study by Kang et al. [11], they indicate that ratios of these measurements might also be helpful. Hence, along with these features, we also compute features in the form of ratios which are T-tau/Aβ1-42, P-tau$_{181}$/Aβ1-42 and P-tau$_{181}$/T-tau for the study.

2.3.4. Striatal Binding Ratio (SBR) from SPECT imaging

SPECT imaging using 123I-Ioflupane was acquired at PPMI imaging centers as per the PPMI imaging protocol (SPECT Manual, http://www.ppmi-info.org/study-design/research-documents-and-sops/) and sent to the Institute of Neurodegenerative Disorders at New Haven, CT, USA for processing and calculation of striatal binding ratios (SBRs). For calculating SBR for each striatal region (left and right caudate, left and right putamen), regions of interest (ROI) were placed on these areas, and the occipital cortex (as the reference region) to extract the count densities for each of these regions (http://www.ppmi-info.org/wp-content/uploads/2013/06/Seibyl-PPMI-MDS-2013-Sydney_sjl.pdf). SBR was then calculated using the formula: SBR = (striatal region/reference region)−1.

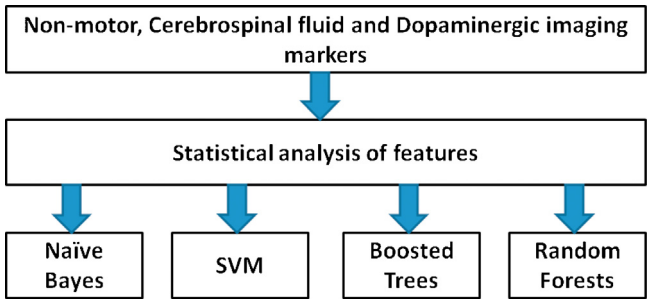

Fig. 1. Flowchart of the proposed analysis.

Table 1
Mean values and standard deviations of features for normal and early PD groups.

Features	Normal ($n_1 = 183$)	Early PD ($n_2 = 401$)	z-statistic	p-value*
RBDSQ score	2.60 ± 2.09	3.26 ± 2.67	−2.42	0.016
UPSIT score	34.08 ± 4.74	22.22 ± 8.23	15.21	≈0
Aβ1-42	376.17 ± 109.65	372.73 ± 99.35	0.66	0.501
α-Syn	2197.9 ± 1085.1	1856.6 ± 795.1	3.45	≈ 0
P-tau$_{181}$	17.98 ± 11.23	15.82 ± 10.20	3.04	0.002
T-tau	52.47 ± 27.10	45.08 ± 18.53	3.05	0.002
T-tau/Aβ1-42	0.16 ± 0.19	0.13 ± 0.07	2.28	0.023
P-tau$_{181}$/Aβ1-42	0.05 ± 0.06	0.04 ± 0.03	2.33	0.020
P-tau$_{181}$/T-tau	0.37 ± 0.19	0.37 ± 0.22	0.34	0.737
SBR$_{RC}$	2.95 ± 0.62	1.99 ± 0.60	14.52	≈ 0
SBR$_{LC}$	3.00 ± 0.65	2.00 ± 0.60	14.65	≈ 0
SBR$_{RP}$	2.15 ± 0.57	0.84 ± 0.36	18.31	≈ 0
SBR$_{LP}$	2.13 ± 0.57	0.81 ± 0.36	18.45	≈ 0

RBDSQ: REM sleep behavior disorder screening questionnaire; UPSIT: University of Pennsylvania smell identification test; Aβ1-42: amyloid beta (1–42); α-Syn: α-synuclein; P-tau$_{181}$: tau phosphorylated at threonine 181; T-tau: total tau; SBR$_{RC}$, SBR$_{LC}$, SBR$_{RP}$, SBR$_{LP}$: Striatal binding ratios of the four striatal regions, namely right caudate, left caudate, right putamen and left putamen.

* p-value «0.01 is shown as ≈ 0. z-statistic and p-value are based on Wilcoxon rank sum test.

2.4. Statistical Analysis of features

A flowchart of the proposed analysis is shown in Fig. 1. We create box plots (Fig. 2) for each feature to visualize its spread and distribution between normal and early PD groups. These box plots show that most of the features are discriminatory. All the features were statistically analyzed for significance using Wilcoxon rank sum test. The means and standard deviations of the features for these two groups are shown in Table 1. We used the Statistics toolbox in MATLAB to perform the test. Only statistically significant features (p-value < 0.05) were used for classification.

2.5. Classification of early PD and healthy normal using automated predictive models

A brief description of classifiers used in the study is given in the online appendix (supplementary document). Let the training data observations be represented as $x_i \in R^m, i = 1, \ldots, n$ where m is the number of features and n is the number of observations used for training, and the class label be represented as y_i such that $y_i \in \{0, 1\}$ where '0' and '1' represent the classes, early PD and healthy normal respectively. We used LIBSVM [28] library for classification

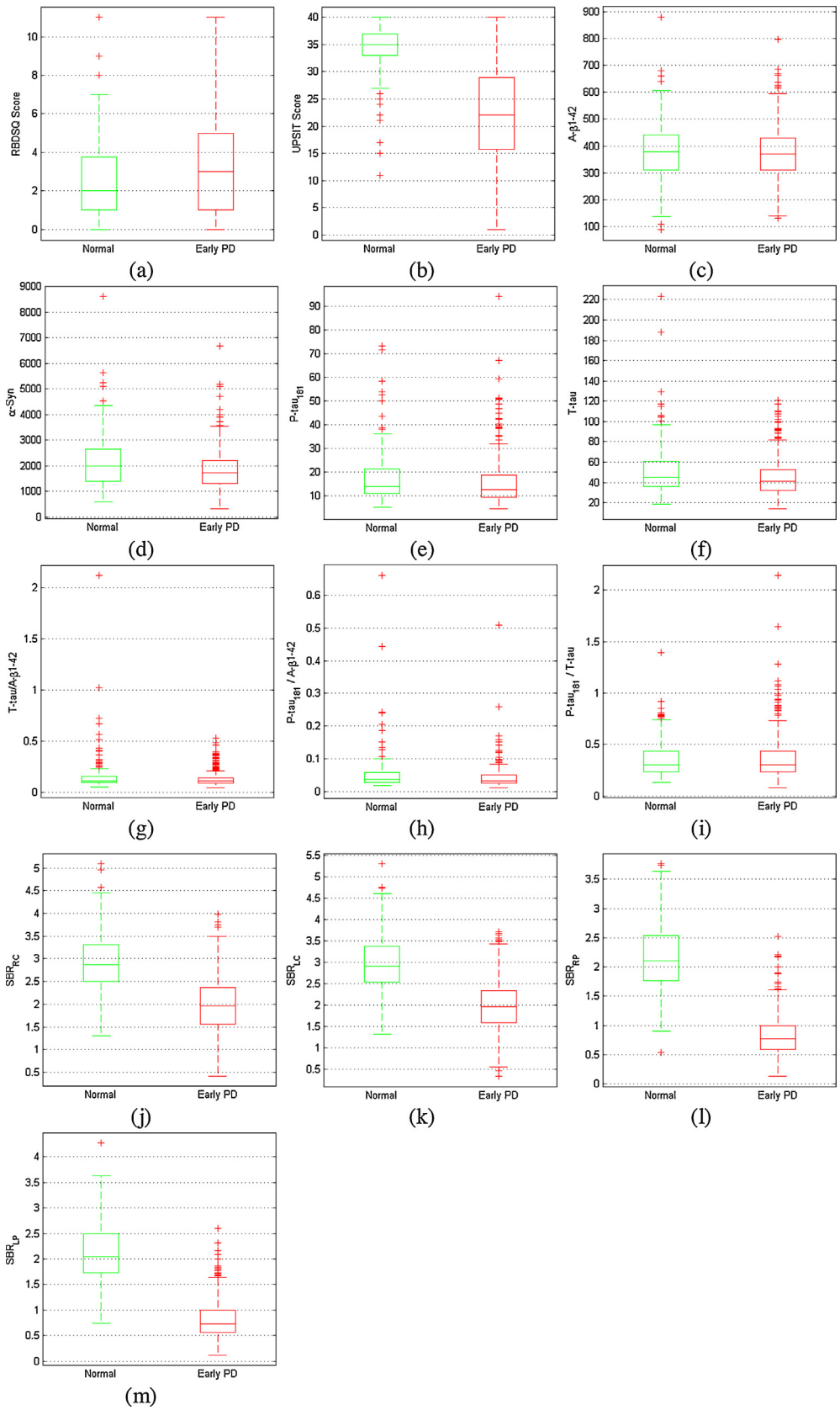

Fig. 2. Box plot of (a) RBDSQ score, (b) UPSIT score, (c) Aβ1-42, (d) α- Syn, (e) P-tau$_{181}$, (f) T-tau, (g) T- tau/Aβ1-42, (h) P-tau$_{181}$/Aβ1-42, (i) P-tau$_{181}$/T-tau, (j) SBR$_{RC}$, (k) SBR$_{LC}$, (l) SBR$_{RP}$, and (m) SBR$_{LP}$, for normal and early PD population.

Table 2
Performance measures for various classifiers used in the study.

Performance measures	Naïve Bayes		Logistic Regression		Boosted Trees		Random Forests		SVM	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
Accuracy (%)	94.67±0.59	93.12 ± 1.49	96.50 ± 0.60	95.63 ± 1.21	100 ± 0	95.08 ± 1.26	99.99 ± 0.02	96.18 ± 1.27	97.14 ± 0.45	96.40 ± 1.08
Sensitivity (%)	94.50±0.68	92.67 ± 2.19	97.38 ± 0.54	96.78 ± 1.55	100 ± 0	96.07 ± 1.72	99.99 ± 0.04	97.55 ± 1.29	97.76 ± 0.42	97.03 ± 1.32
Specificity (%)	95.07 ± 0.97	93.52 ± 3.17	94.56 ± 1.00	93.26 ± 2.82	100 ± 0	92.90 ± 3.74	100 ± 0	93.15 ± 3.60	97.77 ± 1.12	95.01 ± 2.95
AUC (%)	98.66 ± 0.29	96.77 ± 1.33	99.20 ± 0.21	98.66 ± 0.77	100 ± 0	98.23 ± 0.82	100 ± 0	98.40 ± 0.91	99.27 ± 0.16	98.88 ± 0.62

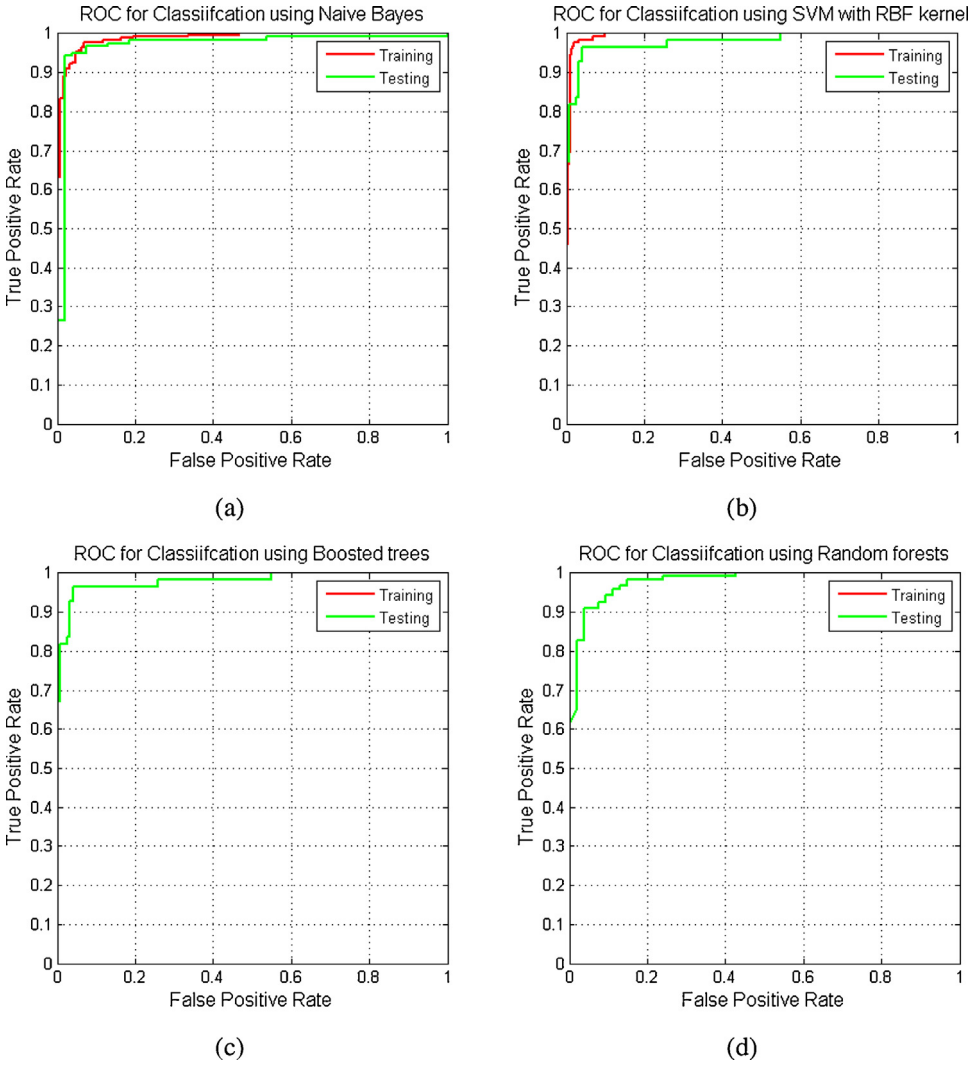

Fig. 3. ROC plots for (a) Naïve Bayes, (b) SVM (c) Boosted Trees, and (d) Random forests classifiers.

using SVM [28,29], Statistics Toolbox in MATLAB for classification using Naïve Bayes [30], Logistic Regression [30], Boosted Trees [31] and Random Forests [32].

About the choice of different classifiers, it is important to bear in mind that there's no one algorithm that's always better than others. This is as per the "No free lunch theorem" [33]. Therefore, one has to try with different set of classifiers and choose the best one. In our study, we have used and compared both, simple (Naïve Bayes and logistic regression) and advanced (SVM, random forests and boosted trees) classifiers. Naïve Bayes and logistic regression come under the category of linear classifiers. Naïve Bayes has the advantage that even if the conditional independence condition of features is not met, it works quite well in many real-world situations. It requires a small amount of training data to estimate the necessary parameters and can be extremely fast compared to more sophisticated methods. Logistic regression is another linear classification technique with the advantage that it gives probabilistic outputs, unlike SVM or more advanced methods. SVM using the radial basis function kernel, boosted trees and random forests are more sophisticated non-linear classification techniques which can easily handle problems which are not linearly-separable. SVM is fast as compared to boosted trees and random forests, but has more parameters to tune and requires normalization of data. Ensemble techniques such as random forests and boosted trees are robust, gives higher accuracy in most cases, requires no normalization, are immune to collinearity, and generates quite good error approximation and are also useful in feature importance ranking. But it is slower than trivial methods like the Naïve Bayes.

We divide the dataset in such a way that 70% is used for training and the rest 30% is used for testing. This partitioning is stratified,

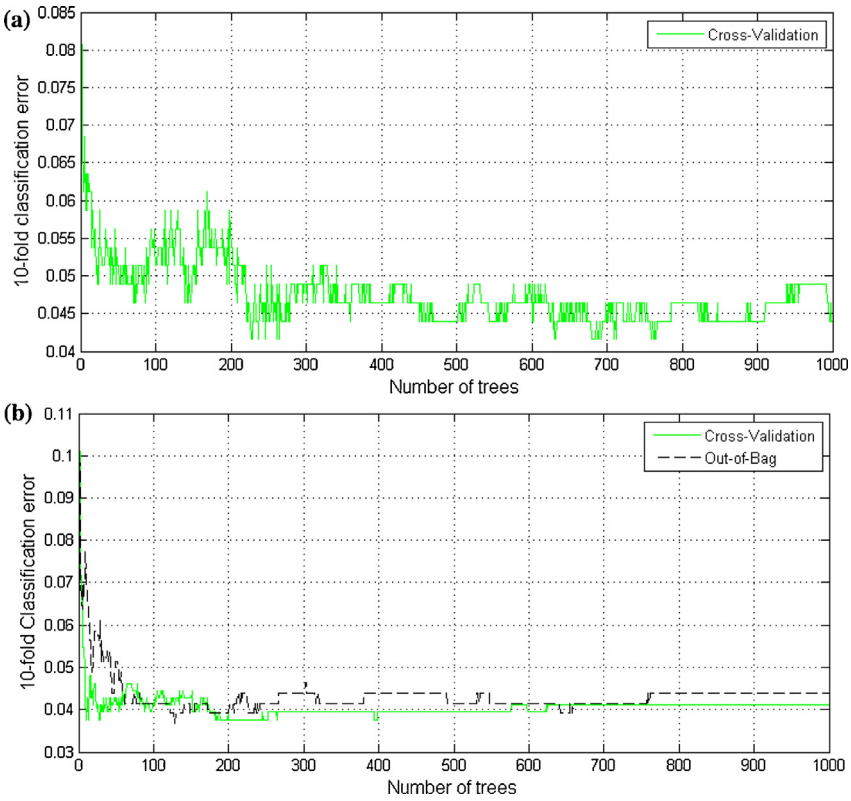

Fig. 4. (a) Plot of 10-fold cross validation (CV) error of Boosted Trees classifier with number of trees (b) plot of 10-fold CV error and out of bag (oob) error of Random forests classifier with number of trees.

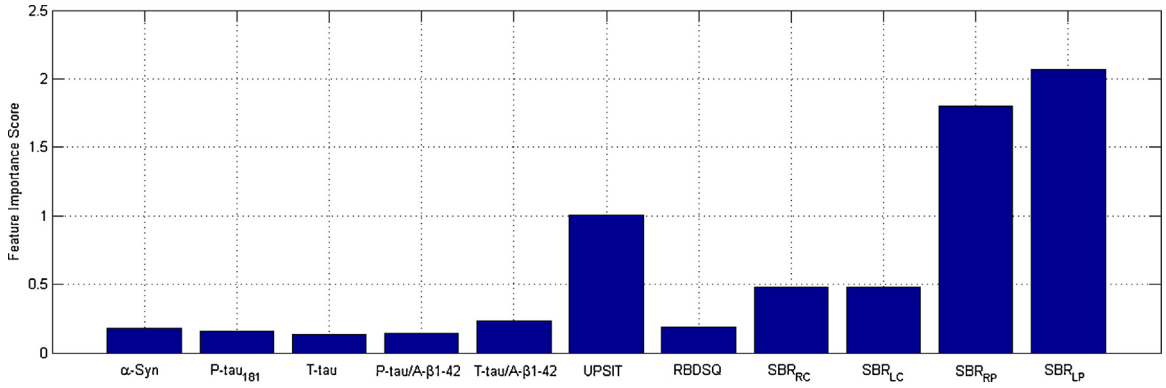

Fig. 5. Plot of importance scores for features. Abbreviations are as given in Table 1. Only statistically significant (Table 1) features which were used for classification are shown.

i.e., both the sets have same class proportions as in the input. And we repeat this procedure 100 times creating different training and testing sets. We compute the accuracy, sensitivity and specificity for these 100 iterations and take the average which is shown in Table 2.

A unique property of 'Random Forests' is that drawing n out of n observations with replacement omits on average 37% of observations for each decision tree. These are '*out-of-bag*' observations, and they can be used to estimate feature importance and predictive power. Out-of-bag estimates of feature importance can be obtained by randomly permuting out-of-bag data across one feature at a time and estimating the increase in the out-of-bag error due to this permutation. Larger the increase, higher the importance of the feature. Out-of-bag error is estimated by comparing out-of-bag predicted responses and the observed responses for all observations used for training.

3. Results and Discussion

Table 1 shows the mean values and standard deviations of features, along with their values of z-statistic and p-value based on Wilcoxon rank sum test, for normal and early PD groups.

Our intention was to use only the statistically significant (p-value < 0.05) features for classification modeling. We observed that 2 out of 13 features (Aβ1-42 and P-tau$_{181}$/T-tau ratio) were not statistically significant and hence, were not used in classification modeling. Our basis of using this testing is from a representative study [34] which suggests that incorporation of external (clinical) information for predictor selection is important for better stability and quality of prognostic models. Kang et al. [11] also observed that these two features had a very low significance value.

The features that were observed to be significant (p-value < 0.05) are: RBDSQ score, UPSIT score, α-Syn, P-tau$_{181}$, T-tau, T-tau/Aβ1-

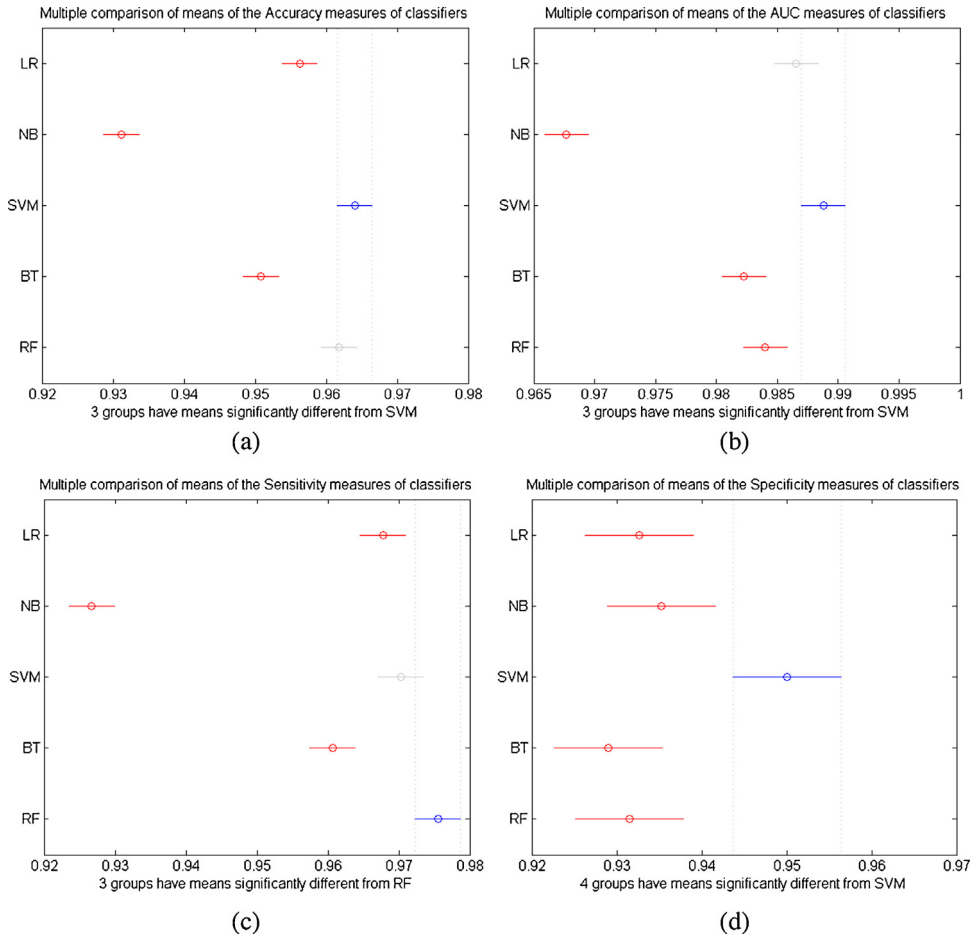

Fig. 6. Statistical comparison of the classifier performance measures using multiple comparison tests (a) Accuracy, (b) Area under ROC (AUC), (c) Sensitivity, and (d) Specificity. Note: LR, NB, SVM, BT, and RF represent Logistic Regression, Naïve Bayes, Support Vector Machine, Boosted Trees and Random Forests, respectively.

42, P-tau$_{181}$/Aβ1-42, SBR$_{RC}$, SBR$_{LC}$, SBR$_{RP}$ and SBR$_{LP}$. These features were used subsequently to perform classification.

Table 2 shows the performances of classifiers used in the study. Fig. 3 shows their ROC plots. The SVM parameters of C and γ for the RBF kernel were obtained using 10-fold cross validation as 512 and 0.0625 respectively. The number of trees in the Boosted trees model and Random forests are selected through 10-fold cross validation. Fig. 4 shows the plot of 10-fold cross validation (CV) error and the number of trees for Boosted trees (Fig. 4(a)) and Random forests (Fig. 4 (b)). Based on these plots, the number of trees for Boosted trees model is chosen as 250, and the number of trees for Random forests is chosen as 100 as this is the point of low CV error and/or out-of-bag (oob) error. The minimum node size for boosted trees in such that each parent node has number of training observations and each child node (leaf) has at least half the number of training observations, and the minimum node size for Random forests is such that each parent node has at least 2 observations and each child node (leaf) has at least 1 observation.

We observe that all the classifiers performed reasonably well with SVM giving the best performance with 96.40% accuracy and 98.88% area under the ROC (AUC). We also observe very low standard deviations (SD) for the performance measures which show that the selection of data for training and testing had low impact on the performance of the classifiers. Fig. 5 shows the plot of feature importance scores as obtained from Random forests. The order of importance of the features is obtained as: SBR$_{LP}$ > SBR$_{RP}$ > UPSIT Score > SBR$_{LC}$ > SBR$_{RC}$ > T-tau/Aβ1–42 > RBDSQ Score > α-Syn > P-tau$_{181}$ > P-tau$_{181}$/Aβ1–42 > T-tau.

In comparison to the related works [11,14–20], all of them used features which are either non-motor features or CSF measurements or dopaminergic imaging markers, but none of them involved combination of premotor features with CSF and imaging markers for the detection of PD. In our study, we use combination of these three entities to perform classification and observe an enhanced accuracy in classifying early PD from healthy normal. This paper extends the work in [11] where they use only CSF measurements and observe that although these measurements showed a significant decrease in PD, they had a low diagnostic utility (AUC <0.8). Although we obtain higher accuracies as compared to related works, it should be noted that the dataset used is different from the other studies.

From Fig. 5, it is observed that among all the features the imaging markers of SBR's for left and right putamen had higher importance than the other features. This indicates that dopaminergic imaging shows higher discriminatory power as compared to other features. The putamenal SBR had higher scores than the caudal SBR which is consistent with previous reports [2,20,35,36]. The feature importance was followed by the UPSIT score and then the CSF measurements of T-tau/Aβ1-42, α-Syn, P-tau$_{181}$, P-tau$_{181}$/Aβ1-42 and T-tau.

A limitation of this study is that the analysis involved early stage (Hoehn and Yahr scale of 1 and 2) PD patients and age-matched healthy normal, and not any premotor PD subjects (subjects who is at risk of PD but not diagnosed as PD due to absence of classic motor symptoms). But the present study is a promising first step towards that long-term goal.

3.1. Statistical Comparison of the classifiers

We carry out one-way ANOVA followed by post hoc tests using Tukey-Kramer method for multiple comparison analysis to determine which pairs of means are significant and which are not [37]. The figure below (Fig. 6) shows various plots obtained. In each graph, the mean of the classifier performance measure is represented by the '*circle*' symbol and an interval (95% confidence interval) around the symbol. Two means are significantly different if their intervals are disjoint, and are not significantly different if their intervals overlap. From these plots, we observe the following: SVM performed the best in accuracy, SVM and Logistic regression performed best in AUC, Random forests and SVM performed best in sensitivity, and SVM performed best in specificity. Therefore, it is explicit that SVM model is giving best results among all and can represent the state-of art for the PD detection problem.

4. Conclusion

Early (or preclinical) diagnosis of PD is important to allow neuroprotective and early management strategies to be administered prior in the disease process. In this work, we use the preclinical markers of non-motor features of RBD and olfactory loss, CSF measurements, and dopaminergic imaging features to perform classification into early PD and healthy normal. We observe that using SVM produced a near perfect classification. The contributions from this study are the following: We perform an extension to the work by Kang et al. where they used CSF measurements and observed a low diagnostic utility. In this work, we use other relevant and significant features corresponding to non-motor and imaging markers and obtain a superior accuracy. We also obtain superior performance in classification as compared to related works. We infer from the study that a combination of non-motor, CSF and dopaminergic imaging measurements have the potential to discriminate early PD from healthy normal and may help in its diagnosis.

Author's Contributions

RP setup the conceptual design, did most of the data analysis and interpretation and wrote the first draft and subsequent revision of the manuscript. SDR and SG collaborated on data analysis and in the writing and revision of the manuscript. PKM collaborated on the conceptual design of the research and in the critical revision of the manuscript for important intellectual content. PPMI provided the necessary data for analysis. All authors have read and approved the final manuscript.

Conflicts of Interest

None

Acknowledgements

PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including Abbott Laboratories, Avid Radiopharmaceuticals, Biogen Idec, Bristol-Myers Squibb, Covance, Élan, GE Healthcare, Genentech, GlaxoSmithKline, Eli Lilly and Company, Merck & Co., Meso Scale Discovery, Pfizer, Hoffmann-La Roche, UCB (Union Chimique Belge), and Piramal.

Summary points

State-of-the-art

- Early detection ofPD is a crucial problem and at present, there are no definite techniques for early diagnosis.
- Research studies have shown that there exist a premotor phase in PD occurring much before the motor phase of PD.
- Premotor features along with imaging features and cerebrospinal fluid markers have shown to be promising features.
- But they were preliminary studies with less number ofsubjects, with lesser accuracy in detection and they did not use a combination of features for PD detection.

Contribution from the Study

- We use a combination of these features for predictive modeling and observe enhanced accuracy from the models (>95%).
- These models have the potential to be used in a clinicalsetting where a clinician can infer information from the models before taking any decision.
- Our study shows that early detection of PD is feasiblethrough multimodal features and machine learning.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ijmedinf.2016.03.001.

References

[1] S. Fahn, Description of Parkinson's disease as a clinical syndrome, Ann N Y Acad Sci 991 (2003) 1–14.
[2] J.L. Cummings, C. Henchcliffe, S. Schaier, T. Simuni, A. Waxman, P. Kemp, The role of dopaminergic imaging in patients with symptoms of dopaminergic system neurodegeneration, Brain 134 (2011) 3146–3166.
[3] S. Sharma, C.S. Moon, A. Khogali, A. Haidous, A. Chabenne, C. Ojo, M. Jelebinkov, Y. Kurdi, M. Ebadi, Biomarkers in Parkinson's disease (recent update), Neurochem Int 63 (2013) 201–229.
[4] K. Stiasny-Kolster, Y. Doerr, J.C. Moller, H. Hoffken, T.M. Behr, W.H. Oertel, G. Mayer, Combination of 'idiopathic' REM sleep behaviour disorder and olfactory dysfunction as possible indicator for alpha-synucleinopathy demonstrated by dopamine transporter FP-CIT-SPECT, Brain 128 (2005) 126–137.
[5] H.W. Berendse, M.M. Ponsen, Diagnosing premotor Parkinson's disease using a two-step approach combining olfactory testing and DAT SPECT imaging, Parkinsonism Relat Disord 15 (Suppl. 3) (2009) S26–30.
[6] M.M. Ponsen, D. Stoffers, J. Booij, B.L. van Eck-Smit, E.C. Wolters, H.W. Berendse, Idiopathic hyposmia as a preclinical sign of Parkinson's disease, Ann Neurol 56 (2004) 173–181.
[7] A. Iranzo, J.L. Molinuevo, J. Santamaría, M. Serradell, M.J. Martí, F. Valldeoriola, E. Tolosa, Rapid-eye-movement sleep behaviour disorder as an early marker for a neurodegenerative disorder: a descriptive study, Lancet Neurol 5 (2006) 572–577.
[8] R.B. Postuma, J.F. Gagnon, M. Vendette, M.L. Fantini, J. Massicotte-Marquez, J. Montplaisir, Quantifying the risk of neurodegenerative disease in idiopathic REM sleep behavior disorder, Neurology 72 (2009) 1296–1300.
[9] K. Marek, D. Jennings, S. Lasch, A. Siderowf, C. Tanner, T. Simuni, C. Coffey, K. Kieburtz, E. Flagg, S. Chowdhury, W. Poewe, B. Mollenhauer, P.-E. Klinik, T. Sherer, M. Frasier, C. Meunier, A. Rudolph, C. Casaceli, J. Seibyl, S. Mendick, N. Schuff, Y. Zhang, A. Toga, K. Crawford, A. Ansbach, P. De Blasio, M. Piovella, J. Trojanowski, L. Shaw, A. Singleton, K. Hawkins, J. Eberling, D. Brooks, D. Russell, L. Leary, S. Factor, B. Sommerfeld, P. Hogarth, E. Pighetti, K. Williams, D. Standaert, S. Guthrie, R. Hauser, H. Delgado, J. Jankovic, C. Hunter, M. Stern, B. Tran, J. Leverenz, M. Baca, S. Frank, C.-A. Thomas, I. Richard, C. Deeley, L. Rees, F. Sprenger, E. Lang, H. Shill, S. Obradov, H. Fernandez, A. Winters, D. Berg, K. Gauss, D. Galasko, D. Fontaine, Z. Mari, M. Gerstenhaber, D. Brooks, S. Malloy, P. Barone, K. Longo, T. Comery, B. Ravina, I. Grachev, K. Gallagher, M. Collins, K.L. Widnell, S. Ostrowizki, P. Fontoura, T. Ho, J. Luthman, M.v.d. Brug, A.D. Reith, P. Taylor, The Parkinson Progression Marker Initiative (PPMI), Prog Neurobiol 95 (2011) 629–635.
[10] M. Shi, J. Bradner, A.M. Hancock, K.A. Chung, J.F. Quinn, E.R. Peskind, D. Galasko, J. Jankovic, C.P. Zabetian, H.M. Kim, J.B. Leverenz, T.J. Montine, C. Ginghina, U.J. Kang, K.C. Cain, Y. Wang, J. Aasly, D. Goldstein, J. Zhang,

Cerebrospinal fluid biomarkers for Parkinson disease diagnosis and progression, Ann Neurol. 69 (2011) 570–580.
[11] J.H. Kang, D.J. Irwin, A.S. Chen-Plotkin, A. Siderowf, C. Caspell, C.S. Coffey, T. Waligórska, P. Taylor, S. Pan, M. Frasier, K. Marek, K. Kieburtz, D. Jennings, T. Simuni, C.M. Tanner, A. Singleton, A.W. Toga, S. Chowdhury, B. Mollenhauer, J.Q. Trojanowski, L.M. Shaw, Parkinson's Progression Markers Initiative, Association of cerebrospinal fluid β-amyloid 1-42, T-tau, P-tau181, and α-synuclein levels with clinical features of drug-naive patients with early Parkinson disease, JAMA Neurol 70 (2013) 1277–1287.
[12] D.J. Brooks, Technology insight: imaging neurodegeneration in Parkinson's disease, Nat Clin Pract Neurol 4 (2008) 267–277.
[13] H.W. Berendse, J. Booij, C.M. Francot, P.L. Bergmans, R. Hijman, J.C. Stoof, E.C. Wolters, Subclinical dopaminergic dysfunction in asymptomatic Parkinson's disease patients' relatives with a decreased sense of smell, Ann Neurol 50 (2001) 34–41.
[14] L. Silveira-Moriyama, J. Carvalho Mde, R. Katzenschlager, A. Petrie, R. Ranvaud, E.R. Barbosa, A.J. Lees, The use of smell identification tests in the diagnosis of Parkinson's disease in Brazil, Mov Disord 23 (2008) 2328–2334.
[15] L. Silveira-Moriyama, A. Petrie, D.R. Williams, A. Evans, R. Katzenschlager, E.R. Barbosa, A.J. Lees, The use of a color coded probability scale to interpret smell tests in suspected parkinsonism, Mov Disord 24 (2009) 1144–1153.
[16] R. Armananzas, C. Bielza, K.R. Chaudhuri, P. Martinez-Martin, P. Larranaga, Unveiling relevant non-motor Parkinson's disease severity symptoms using a machine learning approach, Artificial intelligence in medicine 58 (2013) 195–202.
[17] D.J. Towey, P.G. Bain, K.S. Nijran, Automatic classification of 123I-FP-CIT (DaTSCAN) SPECT images, Nuclear medicine communications 32 (2011) 699–707.
[18] F. Segovia, J.M. Gorriz, J. Ramirez, I. Alvarez, J.M. Jimenez-Hoyuela, S.J. Ortega, Improved parkinsonism diagnosis using a partial least squares based approach, Medical physics 39 (2012) 4395–4403.
[19] I.A. Illan, J.M. Gorrz, J. Ramirez, F. Segovia, J.M. Jimenez-Hoyuela, S.J. Ortega Lozano, Automatic assistance to Parkinson's disease diagnosis in DaTSCAN SPECT imaging, Medical physics 39 (2012) 5971–5980.
[20] R. Prashanth, S. Dutta Roy, P.K. Mandal, S. Ghosh, Automatic classification and prediction models for early Parkinson's disease diagnosis from SPECT imaging, Expert Syst Appl. 41 (2014).
[21] T. Khan, J. Westin, M. Dougherty, Classification of speech intelligibility in Parkinson's disease, Biocybernetics and Biomedical Engineering, 34 34 (2014) 35–45.
[22] I. Mandal, N. Sairam, Accurate telemonitoring of Parkinson's disease diagnosis using robust inference system, International journal of medical informatics 82 (2013) 359–377.
[23] M.M. Hoehn, M.D. Yahr, Parkinsonism: onset, progression and mortality, Neurology 17 (1967) 427–442.
[24] R.L. Doty, P. Shaman, M. Dann, Development of the University of Pennsylvania Smell Identification Test: a standardized microencapsulated test of olfactory function, Physiol Behav 32 (1984) 489–502.
[25] K. Stiasny-Kolster, G. Mayer, S. Schafer, J.C. Moller, M. Heinzel-Gutenbrunner, W.H. Oertel, The REM sleep behavior disorder screening questionnaire—a new diagnostic instrument, Mov Disord 22 (2007) 2386–2393.
[26] T. Nomura, Y. Inoue, T. Kagimura, Y. Uemura, K. Nakashima, Utility of the REM sleep behavior disorder screening questionnaire (RBDSQ) in Parkinson's disease patients, Sleep Med 12 (2011) 711–713.
[27] A. Olsson, H. Vanderstichele, N. Andreasen, G. De Meyer, A. Wallin, B. Holmberg, L. Rosengren, E. Vanmechelen, K. Blennow, Simultaneous measurement of beta-amyloid(1-42), total tau, and phosphorylated tau (Thr181) in cerebrospinal fluid by the xMAP technology, Clin Chem 51 (2005) 336–345.
[28] C.-C. Chang, C.-J. Lin, LIBSVM: A Library for Support Vector Machines, ACM Trans. Intell. Syst. Technol. 2 (2011) 1–27.
[29] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (1995) 273–297.
[30] A.Y. Ng, M.I. Jordan, On Discriminative vs. Generative Classifiers: A comparison of Logistic Regression and Naive Bayes, in: T.G. Dietterich, S. Becker, Z. Ghahramani (Eds.), Advances in Neural Information Processing Systems (NIPS), MIT Press, 2002, pp. 841–848.
[31] Y. Freund, R.E. Schapire, A Short Introduction to Boosting, Journal of Japanese Society for Artificial Intelligence 14 (1999) 771–780.
[32] L. Breiman, Random Forests, Machine Learning 45 (2001) 5–32.
[33] D.H. Wolpert, W.G. Macready, No free lunch theorems for optimization, Evolutionary Computation, IEEE Transactions on 1 (1997) 67–82.
[34] E.W. Steyerberg, M.J. Eijkemans, F.E. Harrell, J.D.F. Habbema, Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets, Statistics in medicine 19 (2000) 1059–1079.
[35] S.J. Kish, K. Shannak, O. Hornykiewicz, Uneven pattern of dopamine loss in the striatum of patients with idiopathic Parkinson's disease. Pathophysiologic and clinical implications, The New England journal of medicine 318 (1988) 876–880.
[36] M.A. Piggott, E.F. Marshall, N. Thomas, S. Lloyd, J.A. Court, E. Jaros, D. Burn, M. Johnson, R.H. Perry, I.G. McKeith, C. Ballard, E.K. Perry, Striatal dopaminergic markers in dementia with Lewy bodies, Alzheimer's and Parkinson's diseases: rostrocaudal distribution, Brain 122 (Pt 8) (1999) 1449–1468.
[37] M.L. McHugh, Multiple comparison analysis testing in ANOVA, Biochemia medica 21 (2011) 203–209.