Prognosis and Diagnosis of Parkinson's Disease Using Multi-Task Learning

Saba Emrani
SAS Institute Inc. R&D
100 SAS Campus Dr.
Cary, North Carolina 27513
semrani@ncsu.edu

Anya McGuirk
SAS Institute Inc. R&D
100 SAS Campus Dr.
Cary, North Carolina 27513
anya.mcguirk@sas.com

Wei Xiao
SAS Institute Inc. R&D
100 SAS Campus Dr.
Cary, North Carolina 27513
wxiao0421@gmail.com

ABSTRACT

Parkinson's disease (PD) is a debilitating neurodegenerative disease excessively affecting millions of patients. Early diagnosis of PD is critical as manifestation of symptoms occur many years after the onset of neurodegenration, when more than 60% of dopaminergic neurons are lost. Since there is no definite diagnosis of PD, the early management of disease is a significant challenge in the field of PD therapeutics. Therefore, identifying valid biomarkers that can characterize the progression of PD has lately received growing attentions in PD research community. In this paper, we employ a multi-task learning regression framework for prediction of Parkinson's disease progression, where each task is the prediction of PD rating scales at one future time point. We then use the model to identify the important biomarkers predictive of disease progression. We adopt a graph regularization approach to capture the relationship between different tasks and penalize large variations of the model at consecutive future time points. We have carried out comprehensive experiments using different categories of measurements at baseline from Parkinson's Progression Markers Initiative (PPMI) database to predict the severity of PD, measured by unified PD rating scale. We use the learned model to identify the biomarkers with significant contribution in prediction of PD progression. Our results confirm some of the important biomarkers identified in existing medical studies, validate some of the biomarkers that have been observed as a potential marker of PD and discover new biomarkers that have not yet been investigated.

CCS CONCEPTS

• **Applied computing → Health care information systems**; • **Computing methodologies → Multi-task learning**; • **Information systems → Data mining**;

KEYWORDS

Parkinson's disease; multi-task learning; biomarker identification; regression

KDD '17, August 13-17, 2017, Halifax, NS, Canada

ACM ISBN 978-1-4503-4887-4/17/08…$15.00
https://doi.org/10.1145/3097983.3098065

1 INTRODUCTION

Parkinson's disease (PD), one of the most common neurodegenerative disorders, is characterized by progressive impairment and deterioration of dopaminergic neurons resulting in dysfunction of movement control and non-motor problems such as depression and anxiety. As many as one million Americans and more than 10 million individuals worldwide are living with Parkinson's disease. Both motor and non-motor manifestations of PD significantly influence the patients and deteriorate their quality of life.

To date, there is no objective medical test to make a certain diagnosis of PD. Instead, the diagnosis is performed using the assessment of motor symptoms such as shaking, rigidity, slowness of movement and postural instability. The motor symptoms however begin to occur in very late stages of the disease when the dopamine concentration is significantly reduced by up to 80% [18]. Starting the treatment at that advanced stage is of little benefit to patients with PD as the degeneration becomes sever. Accordingly, early diagnosis of PD is critical in order to use treatments for managing and delaying disease progression in initial stages. There is a period of at least 5 years and up to 20 years between the start of neurodegeneration and exhibition of clinical motor symptoms. During this period, the patients mainly show non-motor symptoms such as olfactory loss or sleep behavior disorder. These subtle symptoms are not definitive to be used for disease diagnosis, and may be used along with other potential biomarkers for identifying patients at risk of PD. Early diagnosis is therefore a main challenge in the field of PD therapeutics and the absence of a validated biomarker of Parkinson's disease is the major impeding factor in understanding PD progression in order to develop treatments that can delay, prevent or reverse disease progress.

Recognizing the value of such biomarker discovery, Michael J. Fox Foundation (MJFF) for Parkinson's research funded the Parkinson's Progression Markers Initiative (PPMI) in 2011 [17]. The initiative is assisting scientists to establish markers of PD progression by collecting clinical and imaging data and biologic samples from various cohorts. Several clinical rating scales have been also acquired to assess the status of disease in PD patients, such as Unified Parkinson's Disease Rating Scale (UPDRS) [21], Montreal Cognitive Assessment (MOCA) [19] and Scales for Outcomes in Parkinson's disease for Autonomic Symptoms (SCOPA-AUT) [33]. Since neurodegerenation of PD advances many years prior to manifestation of symptoms and the treatment is more effective in early stages, it is crucial to achieve two key objectives. First, we aim to predict the progression of PD at multiple future time points measured by clinical rating scales such as UPDRS. We next focus on identifying features that are most predictive of PD progression to investigate

existing evidence for biomarkers and to identify new clinical, biological and brain imaging markers.

We employ a multi-task learning regression framework to predict PD progression for up to 4.5 years and to identify important predictive biomarkers from the learned model. We model the prediction of disease severity quantified by clinical rating scales at a series of time points as a multi-task learning problem, where each task is prediction of a clinical scale at one time point. The conventional machine learning frameworks consider only one learner seeking to solve a single task, independent of the others. In numerous applications, however, there are multiple tasks labeling the same data instances differently, in different times or conditions. When the tasks are related, the information learned from each task can be used to enhance the learning of other tasks. It is therefore beneficial to learn relevant tasks together simultaneously, as opposed to learning each task independently. Multi-task learning uses the intrinsic relations between multiple tasks to improve the generalization performance. In this study, a patient's clinical data at different consecutive visits cannot be assumed to be fully independent. Accordingly, the multi-task learning can benefit prediction of disease status in all visits by leveraging their relatedness and shared information across the visits. We use convex fused sparse group Lasso (cSFGL) formulation that concurrently chooses a shared set of features for all visits and a particular set of features for each distinct visit, and at the same time uses fused lasso to integrate the temporal smoothness [32, 38].

Extensive experimental results have been carried out using data from PPMI database. We accurately predict the disease progression over 11 visits that span a course of 4.5 years. We substantiate the effectiveness of multi-task learning framework for PD prognosis compared to single learner methods. We also fully exploit the method to identify important diagnostic biomakers. Some markers selected by our experimental results are consistent with findings from medical and clinical studies. We validate these established biomarkers and discover new biomarkers that have not been recognized to date. The biomarkers discovered in this study would help substantially expedite Parkinson's disease therapeutics research towards early diagnosis and development of effective treatments. We now describe related research in this important area.

2 RELATED WORK

Related research in a number of areas is relevant to this study, including (i) identifying biomarkers of PD; (ii) multi-task learning for modeling disease progression; and (iii) methods for capturing the relatedness of the tasks.

Non-motor symptoms such as sleep disorder and olfactory loss, in conjunction with other important biomarkers including Cerebrospinal fluid (CSF) and brain imaging scan data from PPMI are used in [25] to differentiate two groups of 183 healthy subjects and 401 patients with early Parkinson. Naive Bayes, Support Vector Machine (SVM), Boosted Trees and Random Forests classifiers are employed for classification and the best performance is achieved by SVM. An enhanced probabilistic neural network is used in [10] with non-motor symptoms, clinical and behavioral assessments and brain imaging features for differentiating PD patients from subjects whose scans do not exhibit evidence of dopaminergic deficit. CSF measurements are extensively investigated for identification of neurodegenerative diseases. These biomarkers have recently been shown to hold promise in PD diagnosis although these investigations are in very early stages. A study with 63 early PD patients and 39 healthy normal subjects from PPMI database used multivariate regression analysis and showed that the amount of particular CSF biomarkers is slightly but significantly lower in patients with PD compared to normal control [14]. The identification capability of these biomarkers however were demonstrated to be weak with less than 80% accuracy and as a result they need to be used along with other valid biomarkers to enhance the diagnosis. The mentioned study leaves further analysis of predictive performance of CSF biomarkers and testing them on all PPMI subjects to future work. Another study presented in [15] also focused on CSF measurements in 660 PPMI subjects at baseline, and used the correlation of these features with the clinical assessments to identify a subset of CSF features with significantly different levels in PD patients compared to healthy control. The existing work in PD biomarker discovery including the ones mentioned above are all focused on classifying PD patients from healthy control subjects. They focus on identifying biomarkers for differentiating the two groups and are not capable of predicting the progression of disease. Moreover, these studies do not investigate many other available candidate biomarkers such as RNA, plasma and serum measurements.

Experimental results have shown advantages of multi-task learning compared to single independent learners in problems involving related tasks [1, 4, 9]. Multi-task learning has been effectively used in a wide variety of applications including object location and recognition in image processing [3], speech classification [22], data integration from different web directories [26], identification of handwritten digits [26] and multiple microarray data integration in bioinformatics [34]. Multi-task learning approaches have been lately proposed for modeling the Alzheimer's disease progression [38].

There are various ways to impose the relatedness of tasks in order to incorporate the insight from task connections in joint learning. The first category of multi-task learning approaches presumes that all tasks are related and the knowledge in each task is shared with all others. In this category, all tasks are either considered to have models close to each other or share a joint set of features [7, 20]. In the second category of approaches, tasks are arranged in clusters where tasks in each cluster have models closer to each others. Task clustering is enforced in [1] by using several clusters of similar tasks and taking a mixture of Gaussians as a prior instead of a single Gaussian. A task clustering regularization is used in [6] to incorporate cluster information in multi-task learning methods as an extension to kernel learning techniques. A nonparametric hierarchical Bayesian model is used in [35] to automatically identify task clusters. Third category of multi task learning methods is graph structured techniques where the task relationship is captured using a graph [6]. Each task is represented by a node, and two nodes are connected with an edge if the corresponding tasks are related to each other. The similarity degree of a pair of tasks is then represented by the weight of the linking edge. In this study, we use a graph to model the task relatedness and integrate our knowledge of the nature of this relationship into multi-task learning framework.

3 DATASET

Data used in this paper is obtained from PPMI database at www.ppmi-info.org/data at the time of this study. PPMI is a landmark, large-scale, comprehensive longitudinal study that is currently being conducted on PD patients as well as healthy normal subjects. The initiative follows more than 1,000 participants for up to 8 years, at a network of 33 clinical sites around the world. PPMI is the first significant observational study created to identify and validate biomarkers for prediction of PD progression [17]. The goal of this initiative is using these biomarkers to improve the understanding of PD progression and provide essential means for controlling PD in therapeutic trials. PD patients in the PPMI study are de novo patients, recently diagnosed and unmedicated. The dataset contains an inclusive set of clinical and behavioral assessments, brain neuroimaging scans and biospecimens including CSF, DNA, RNA, plasma, serum, urine and cell lines. Baseline data is gathered at the beginning of the study when the patient performs the initial screening. Currently, PPMI dataset contains measurements from 12 follow-up visits after the baseline. We use data from 11 visits carried out during 4.5 years and do not use data from 12^{th} visit due to the its small sample size.

4 PREDICTION AND PROGRESSION BIOMARKERS

We now present the problem, summarize features and potential target scores measured from patients, and review the prediction model.

4.1 Problem Description

We have access to clinical motor and non-motor assessments, brain imaging scans, biospecimen collection and lab samples obtained from 400 patients with early Parkinson's disease and 200 healthy control subjects. These measurements and tests are acquired repeatedly over a 5-year interval. The disease status is also assessed multiple times in the recurrent visits using clinical rating scales. The first objective is to predict the progression of disease by forecasting the cognitive scores using only the measurements at the baseline. The second objective is identifying one or more biomarkers of Parkinson's disease progression as crucial means to understand the disease and develop new and more effective treatments. Prediction of cognitive scores at each future visit can be performed using a regression model and multi-task learning can be exploited for prediction of the disease rating scales at multiple future visits. By using multi-task learning, the prediction performance benefits from the important fact that a subject's disease status at different time points are not independent. This temporal information between tasks will be integrated in the model to enhance the disease prediction performance. We then use the model learned from multi-task regression to find the features most predictive of the progression and their corresponding importance.

4.2 Features and Targets

We now describe the features extracted from PPMI database. We group available features into three main categories: (i) clinical assessments, (ii) biological specimen, and (iii) brain imaging scans. We also discuss in this section, the scores for assessing the status or severity of PD that are potential targets of the multi-task learning.

4.2.1 Clinical Assessments. This class contains features obtained from questionnaires answered by patients or their health care professionals who perform the examinations on the patients.

Smell Identification Test: Severe olfactory disorders such as impairment in detecting, distinguishing and recognizing different smells are observed at early stages of Parkinson's disease [23]. The study in [23] shows that idiopathic olfactory dysfunction is associated with a 10% higher risk of developing PD. The University of Pennsylvania Smell Identification Test (UPSIT) is an accurate, reliable and long-established smell identification assessment [5]. It includes four booklets each containing ten pages where each page includes a distinct odor in a plastic microcapsule. The participant releases the odor by scratching the container and then identifies the smell by selecting one of the four choices that best represents the smell. The maximum possible score is 40 when all the smells are recognized correctly. The UPSIT feature will be a number between 0 and 40 for each subject.

Sleep Behavior Disorder Screening: Another non-motor symptom of PD is Rapid Eye Movement (REM) sleep Behavior Disorder (RBD) that generally precedes the neurodegenrative diseases [13]. RBD is a sleep disruption in which the patient physically and suddenly acts on vivid, aggressive and unpleasant action dreams. The dream-enacting behaviors include talking, shouting, kicking, punching, sitting, jumping from bed or other violent limb movements. A clinical study has been performed over 12 years on 93 subjects with RBD that shows the patients with REM sleep disorder are generally in increased risk of developing neurodegenerative disroders [24]. In this study, 28% of the patients developed neurodegenerative diseases out of which 15% had Parkinson's disease. RBD Screening Questionnaire (RBDSQ) is a particular assessment questionnaire for RBD that evaluates the most significant clinical features of RBD. It has been extensively studied and validated to be a useful tool for the screening of RBD in PD patients due to its high sensitivity and specificity [31]. RBDSQ includes 12 questions with 'yes'/'no' answers. We assign score 1 to 'yes' and score 0 to 'no' answers. The RBDSQ feature will be an integer between 0 and 12 for each patient, where higher scores signify higher probability of RBD disorder.

4.2.2 Biologic Specimens. Biologic specimens used in this study include cerebrospinal fluid (CSF), plasma, serum and RNA. CSF is a fluid surrounding the brain and spinal cord that indicates pathological condition of the central nervous system. CSF measurement is more accessible and less costly than brain imaging. Four CSF biomarkers of amyloid beta peptide 1-42 ($A\beta$1-42), total tau (T-tau), tau phosphorylated at threonine 181 (P-tau181) and $\alpha-$synuclein ($\alpha-$Syn) are extensively studied and identified as sensitive and specific biomarkers for early diagnostic differentiation of patients with Alzheimer's disease (AD) from healthy individuals [29]. The first study of CSF biomarkers in PPMI study subjects revealed that ratios of these biomarkers might also be important [14]. In this study, we therefore consider seven CSF features including $A\beta$1-42, T-tau, P-tau181, $\alpha-$Syn as well as ratios T-tau/$A\beta$1-42, P-tau181/$A\beta$1-42 and P-tau181/T-tau. Additionally, plasma measurements from 154 PD patients from the PPMI study is analyzed for epidermal growth factor (EGF) and Apolipoprotein A1 (ApoA1). Lipid fats in the blood

including total cholesterol, HDL (high-density lipoprotein), LDL (low-density lipoprotein) and triglycerides were also measured, resulting in 6 plasma features. We also consider the expression of 19 RNA biomarkers that had enough observations as well as one serum measurement, namely serum Insulin-like Growth Factor 1 (IFG-1).

4.2.3 Brain Imaging. SPECT (Single Photon Emission Computerized Tomography) scan is a type of nuclear imaging that uses small amounts of a radioactive injection and multiple cameras to provide rapid dynamic 3D acquisitions. DaTscan is an imaging agent for use with SPECT to detect dopamine transporters (DaT) in the brain for patients with Parkinson's disease. Although DaTscan cannot diagnose PD, it is used to help confirm a diagnosis. Imaging studies have shown a distinct decrease in dopamine transporter binding in subregions of the striatum in PD patients [2]. DatScan SPECT imaging data acquired at PPMI are processed to compute binding ratios for each striatal region (left and right caudate, left and right putamen). Striatal binding ratios(SBR) is calculated by positioning the regions of interest (ROI) on the four striatal regions mentioned above as well as on a reference area to derive the count densities for each of the regions. SBR is then calculated using SBR = (count density of striatal region/count density of reference region)-1. We use four SBR features calculated in left and right caudate and left and right putamen.

4.2.4 Target Scores. **MDS-UPDRS:** Movement Disorder Society revision of the United Parkinson's Disease Rating Scale (MDS-UPDRS) is a universal scale for Parkinson's symptoms created to comprehensively assess and document the examination of patients with PD. It includes three sets of clinical examinations. Part I consist of 13 questions answered by the patient, assessing non-motor experiences of their daily life such as cognitive impairment, depression and anxiety. Part II consist of 13 questions about motor experiences of patient's daily living such as difficulties with speech, chewing, handwriting and walking. Part III consists of 18 questions that are answered by an examiner after assessing patient's motor functions such as observing the patient's facial expressions. For each item, a number between 0 and 4 is assigned as a rate to indicate the severity of the corresponding symptom. We use the total MDS-UPDRS score that is summation of the three scores as a potential target.

Montreal Cognitive Assessment (MoCA) consists of 26 items that assess cognitive function such as attention, concentration, memory, language, conceptual thinking and orientation. This test is widely used as a cognitive screening test in Parkinson's disease.

SCOPA-AUT: The Scales for Outcomes in Parkinson's disease for Autonomic Symptoms (SCOPA- AUT) consists of 21 items that assess patient's autonomic problems such as thermoregulatory or urinary dysfunction.

4.3 Prediction Model

The prediction model uses the features described in the previous section, measured on each patient, to make prediction about the status of each patient's Parkinson's disease in the future. We use an appropriate rating score of Parkinson's disease as targets that are acquired from the patients repeatedly at multiple time points. By viewing the prediction of disease score at a single time point as one regression task, we formulate the progression of PD rating score over several time points in the future as a multi-task regression problem. Each task is prediction of the target for a group of patients at each future time point. Consider a total number of k tasks representing k time points. For each task $i \in \{1, \cdots, k\}$, assume that there are p features and n observations or patients. Let $X \in \mathbb{R}^{n \times p}$ be the input data or observation matrix including features measured at the baseline. The corresponding target matrix containing PD rating scores of n patients at k time point is denoted by $Y \in \mathbb{R}^{n \times k}$. Let $W = [w_1, \cdots, w_k] \in \mathbb{R}^{p \times k}$ be the model. A linear approach for learning W is by solving the following regularization problem: $\min_W \|XW - Y\|_F^2 + \theta_1 \|W\|_F^2$, where θ_1 is the regularization parameter controlling the generalization error and $\|.\|_F$ is the Frobenius norm. The regression problem above, known as *ridge regression* does not take into account the relatedness of different tasks and treats them as independent. However, the status of a patient's PD at different consecutive time points are not independent of each other. To take into account the relatedness of PD scores at various consecutive times, we add a graph regularization term in the regression model that represents the relationships between tasks in the multi-task learning formulation [37]. The relationship between k tasks is represented by a graph, where each task is a node, and two nodes are connected if the corresponding tasks are related. To capture the connections, we denote the set of edges in the graph by E. The graph is encoded using a $\|E\| \times k$ matrix R, where $\|E\|$ is the number of edges. Let e_j^i denote edge i connected to node j, which is zero if the node is not connected to the edge. If nodes x and y are connected via edge e^i, then e_x^i and e_y^i are set to 1 and -1, respectively. e^i will then form the i^{th} row of the matrix R, which is defined as $R = \left[e^1 \quad e^2 \quad \cdots \quad e^{\|E\|}\right]^T$.

Figure 1: The graph structure for capturing the task relationship in Parkinson's disease. Each node represents an individual task and a pair of nodes are connected when the two corresponding tasks are related. The R matrix for this graph is expressed in Equation (1).

One way of assigning the task relationships is based on the prior information about the specific problem. In this study, since it is established that PD is a progressive disease, it is reasonable to consider that the difference between predictions at consecutive time points is relatively small as the disease deteriorates. This prior information based on temporal smoothness has been effective for analyzing Alzheimer's disease progression which is another neurodegenartive disorder [39]. For the case in this study where each task is related to its subsequent tasks, we therefore use the graph structure illustrated in Figure 1. Using this graph and the definition of R matrix, the corresponding R can be determined as a $(k-1) \times k$

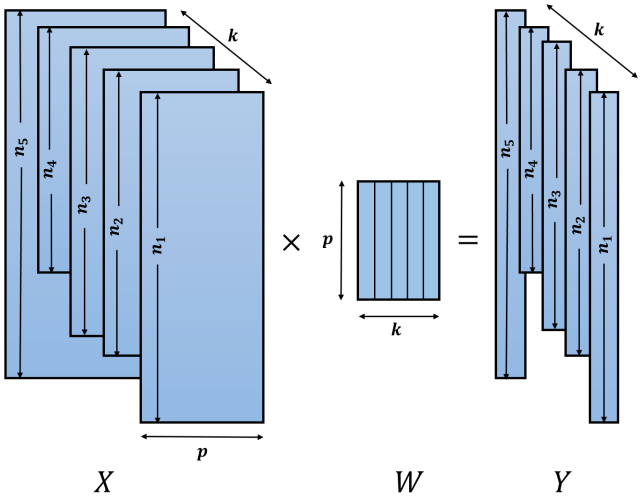

Figure 2: Illustration of the capability of multi-task regression in handling missing target values

matrix shown below

$$R = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & 0 & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 1 & -1 \end{bmatrix}, \quad (1)$$

where $R_{i,i} = 1$ and $R_{i,i+1} = -1$. Using the R matrix above, the additional penalty term $\|RW^T\|_1$ can be defined as $\|RW^T\|_1 = \sum_{i=1}^{k-1} |w_i - w_{i+1}|$. This term in the regularization will penalize large variations of predictions at consecutive time points to help keep the temporal smoothness. Moreover, it is established in the past few years that a small subgroup of biomarkers contribute to PD progression and different biomarkers may be associated to various stages of the disease [28]. In addition to the graph regularization term, we therefore use a group Lasso penalty defined as

$$\|W\|_{2,1} = \sum_{i=1}^{p} \sqrt{\sum_{j=1}^{k} W_{ij}^2} \quad (2)$$

to incorporate features common among different tasks. We also integrate features specific to each particular task using Lasso penalty. The final optimization problem can therefore be represented as:

$$\min_W \|XW - Y\|_F^2 + \theta_1 \|RW^T\|_1 + \theta_2 \|W\|_{2,1} + \theta_3 \|W\|_1. \quad (3)$$

This formulation is called convex fused sparse group Lasso (cFSGL) [38]. The optimization problem above is solved using accelerated gradient method (AGM). We applied the technique included in the MALSAR package [37] to efficiently solve the optimization. For comparison, we also implement temporal group Lasso (TGL) proposed in [39] expressed in the following formulation:

$$\min_W \|XW - Y\|_F^2 + \theta_1 \|RW^T\|_2 + \theta_2 \|W\|_{2,1} + \theta_3 \|W\|_2. \quad (4)$$

Table 1: Complete list of features analyzed in this study

Type	Features
Biologic specimens	Cerebral Spinal Fluid (CSF): $A\beta$1-42, T-tau, P-tau181, α−Syn, T-tau/$A\beta$1-42, P-tau181/$A\beta$1-42, P-tau181/T-tau RNA: DHPR, DJ-1, FBXO7-001, FBXO7-005, FBXO7-007, FBXO7-008, FBXO7-010, GLT25D1, GUSB, MON1B, RPL13, SNCA-007, SNCA-3UTR-1, SNCA-3UTR-2, SNCA-E3E4, SNCA-E4E6, SRCAP, UBC, ZNF746 Plasma: Total Cholesterol, ApoA1, EGF, HDL, LDL, , Triglycerides Serum: IFG-1
Clinical assessments	UPSIT, RBDSQ, MDS-UPDRS
Brain imaging	SBR caudate right, SBR putamen right, SBR caudate left, SBR putamen right

4.4 Missing Targets

The PD rating scores for many patients are missing at some visits, resulting in missing values in target matrix Y. One solution is to completely remove from the experiment each patient that has missing target values in at least one visit. This approach results in a considerably decreased number of observations. However, the multi-task learning formulation described in Section 4.3 is capable of handling missing values in targets since it does not necessarily require the tasks to have the same number of observations. The same matrix of features measured at the baseline X is used to predict the target values at all future time points. Hence, the samples with missing target values in one task can be removed only from the feature matrix used to estimate target values of the corresponding task. In other words, if a patient's target score is missing at one visit, the features measured on that patient in the baseline will be removed from only the feature matrix of that task, without causing problem for W matrix. Figure 2 demonstrates how multi-task regression can handle missing target values.

5 EXPERIMENTAL RESULTS

We now present the findings of our experiments. We first report the disease prediction or prognosis performance in Section 5.1. We then investigate the disease diagnosis by inspecting the model to understand the contributions that each feature makes towards prediction of the disease status, and the performance of different feature categories.

First, we analyze the capability of potential targets described in Section 4.2 including MDS-UPRDS, MoCA and SCOPA-AUT for assessment of Parkinson's disease status. There is no valid ground truth information about the severity of disease. However, we need to evaluate if these scores are correct measures of disease status. Since PPMI dataset includes definite diagnosis indicating PD and control patients, we use that information in order to validate which score can be used for measuring disease severity. If a score is not useful in differentiating between patients with PD and healthy control subjects, it cannot be used as a measure of disease condition.

Table 2: Number of observations (patients) in different visits for MDS-UPDRS using CSF features and all features (CSF+A) together with the data dimensionality (Dim) denoting the number of features.

Source	V01	V02	V03	V04	V05	V06
CSF	373	274	324	352	356	349
CSF+A	99	101	92	99	97	85
Source	V07	V08	V09	V10	V11	Dim
CSF	351	347	285	208	128	7
CSF+A	91	94	91	79	46	40

Figure 3 shows the histogram with overlaid normal and kernel densities and the box plot for each of the potential targets discussed in Section 4.2 for the control and PD subjects groups. In the box plots, the symbol marker represents the mean score, and the left and right edges of the box represent the first and third quartiles. The ends of the whiskers show the most extreme score value not considered outliers and a point is considered an outlier if it is more than $1.5 * (q_3 - q_1)$ above q_3 or below q_1 percentile, where q_1 and q_3 are the 25th and 75th percentiles, respectively. With p-value of $p < 0.0001$, MDS-UPDRS is significantly different for patients with PD and control group. However, as it is clear in Figure 3, MoCA and SCOPA-AUT are not significantly different for control and PD with large overlaps between the histograms for the two groups. Accordingly, we only use MDS-UPDRS score as target for prediction of PD status.

The complete list of features used in this study is shown in Table 1. We include the MDS-UPDRS score measured at baseline in the feature matrix since it is correlated with the future MDS-UPDRS scores and can help their prediction. We use CSF and combination of CSF with other features (CSF+A) at baseline to predict MDS-UPDRS score for 11 future visits happening during 4.5 years after the initial screening visit. In the X matrix, we remove samples with at least one missing feature. As a result, the feature set includes the intersection of available samples, such that the merged feature matrix has no missing values. With the described feature engineering process, the number of observations available for predicting MDS-UPDRS using CSF and CSF+A at all 11 visits are shown in Table 2.

5.1 Prognosis: Prediction Performance

In this Section, we apply the multi-task learning models including convex fused sparse group Lasso (cFSGL) presented in Equation (3) and temporal group Lasso (TGL) in Equation (4) to the features outlined in the Section 4.2, measured from each patient, to make predictions about the status of PD in each patient's future timelines. We also include the results from nonconvex fused sparse group Lasso (nFSGL) formulation from [38] for a complete comparison. We compare the results with two single learner methods including ridge regression and Lasso on the prediction of MDS-UPDRS using different feature combinations, namely CSF and CCSF+A. For each feature combination, we randomly split the data into training and testing sets. We use train to test ratios of 9:1 for CSF and 7:3 for CSF+A datasets. The reason for using a larger test to train ratio in CSF+A dataset is fairly small number of observations in this dataset after removing missing features, as shown in Table (2). We use

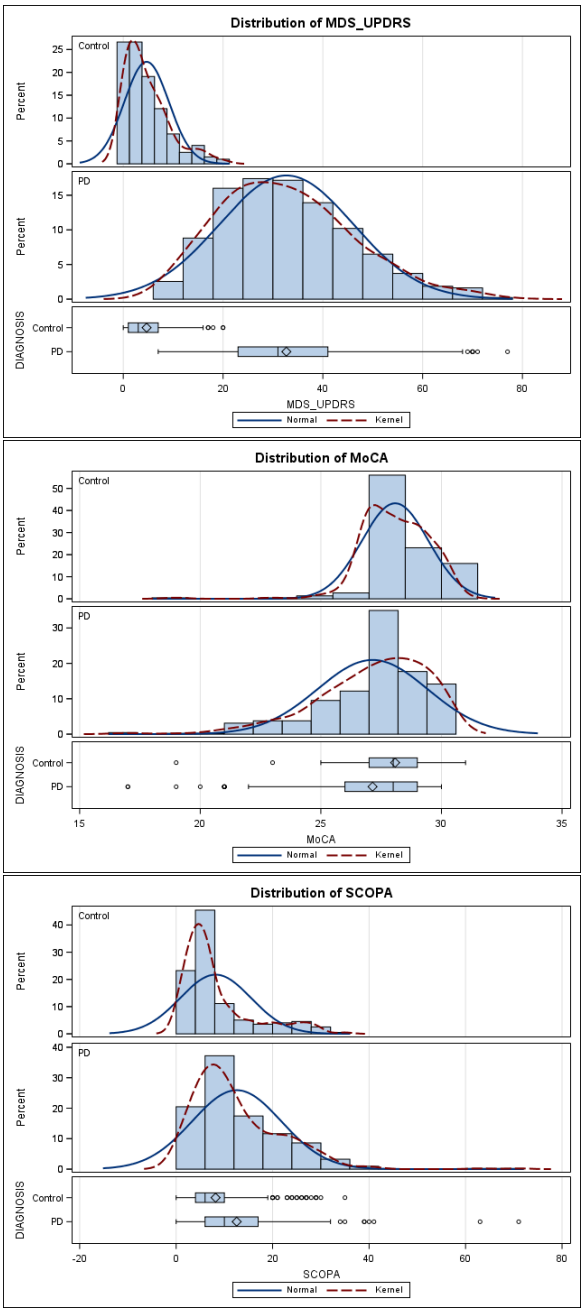

Figure 3: The histogram of MDS-UPDRS, MoCA and SCOPA-AUT values for patients with PD and control group. Overlaid normal and kernel densities and the box plot are also shown for each of the potential targets for the control and PD groups. MDS-UPDRS is significantly different for PD and control ($p < 0.001$). MoCA and SCOPA-AUT are not significantly different in PD and control and cannot be used as target measurements of disease status.

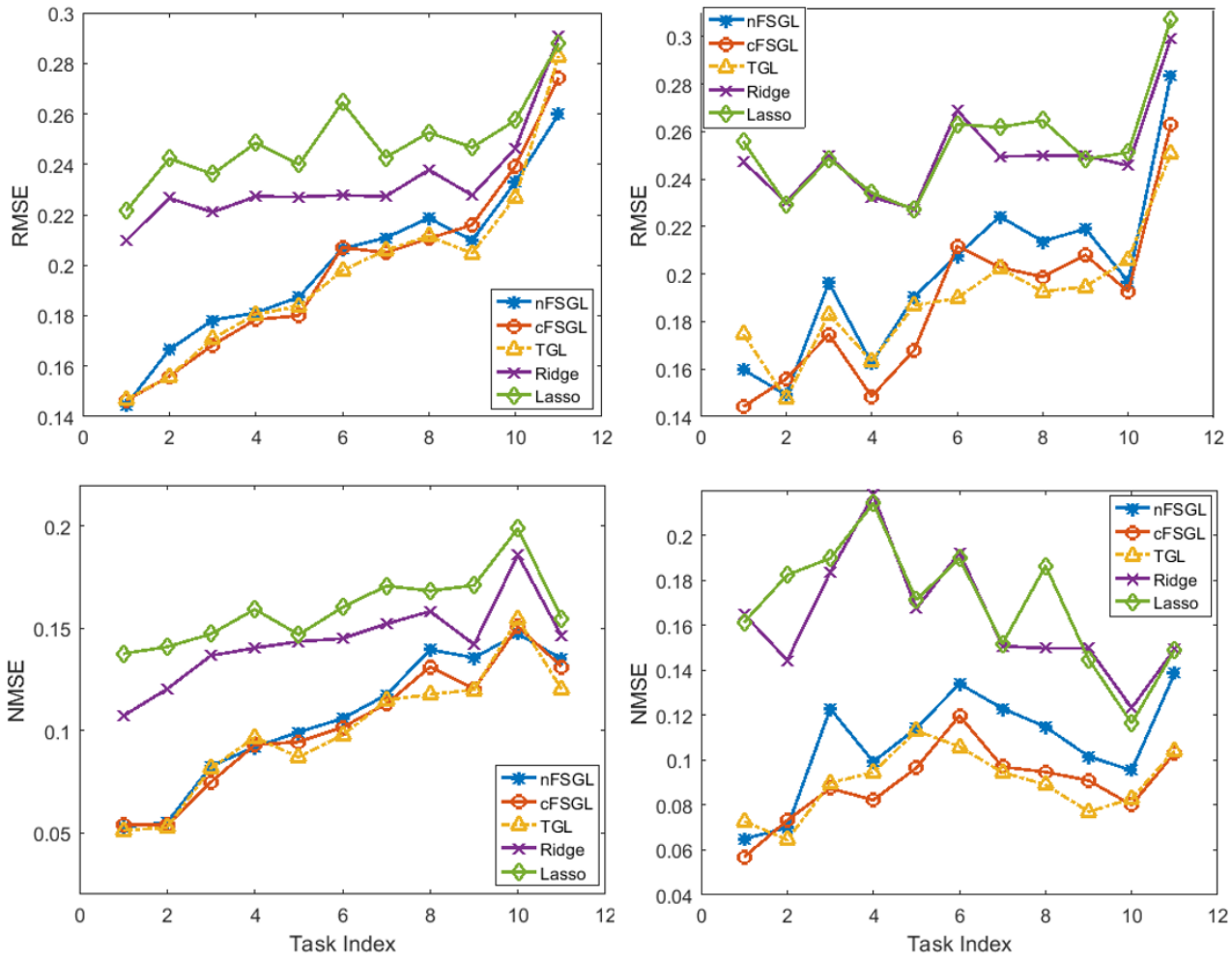

Figure 4: Comparison of different multi-task learning approaches (TGL, cFSGL, nSFGL) with single learner approaches (ridge regression and Lasso) on longitudinal MDS-UPDRS prediction using different feature combinations. The left and right columns show prediction results using only CSF feature and all features (CSF+A), respectively. The top and bottom row show RMSE and NMSE, respectively. Each task represents a visit and 11 visits are performed over 4.5 years.

Table 3: Comparison of TGL, cFSGL and nSFGL with ridge and Lasso on MDS-UPDRS prediction using CSF and CSF+A features in terms of average and standard deviation of RMSE and NMSE over all tasks.

CSF	Ridge	Lasso	TGL	cFSGL	nFSGL
RMSE	0.245±0.046	0.259±0.038	0.210± 0.057	0.210± 0.054	0.212±0.052
NMSE	0.148± 0.025	0.161±0.017	0.103±0.032	0.106±0.035	0.110±0.035
CSF+A	Ridge	Lasso	TGL	cFSGL	nFSGL
RMSE	0.259± 0.035	0.262±0.035	0.195 ± 0.030	0.192 ±0.036	0.208 ±0.045
NMSE	0.158±0.029	0.165±0.029	0.087 ± 0.016	0.086 ±0.018	0.105 ± 0.023

5-fold cross validation to select the model parameters θ_1, θ_2 and θ_3 in each of the Equations (3) and (4) in the training data. Normalized Root Mean Square Error (RMSE) and Normalized Mean Squared Error (NMSE) are employed for performance evaluation as widely used in multi-task learning studies [36, 38]. The experimental results for each of the 11 tasks on CSF and CSF+A using different methods mentioned above are included in Figure 4. Additionally, the results for overall performance of the prediction averaged across all visits are summarized in Table 3, where the mean and standard

deviation of RMSE and NMSE are obtained using 100 repetitions of experiments on various splits of data into train and test sets.

We have the following observations regarding the prediction of PD progression: Our results show that we can accurately predict the future status of the Parkinson's disease for up to 4.5 years using multi-task learning. We also show that all the tested multi-task learning approaches (TGL, cFGL and nFSGL) significantly outperform single learner methods (ridge and Lasso) in each task as illustrated in Figure 4 and over all tasks as shown in Table 3. Our results also demonstrate that cFSGL has the best performance among the implemented methods for prediction of PD progression. Moreover, all three multi task learning methods perform better in CSF+A dataset compared to CSF dataset, while single learner methods show higher error in CSF+A. The reason is that the feature space is larger and the number of observations is smaller in CSF+A compared to CSF (see Table 2) and multi-task learning expands the size of sample set by simultaneously learning multiple tasks. On the other hand, ridge and lasso approach each task independent of the others and suffer from smaller number of observations. These results show that the future status of PD can be predicted more accurately by including all the features and using multi-task learning even with a smaller sample size. Therefore, it is better to have complete measurements from less patients in the clinical trail than only collecting cerebral spinal fluid from much larger number of patients.

5.2 Diagnosis: Feature Contributions

In addition to the progression prediction, we are interested in identifying the biomarkers that are most predictive of the status of Parkinson's disease. We first experiment with different sets of features separately, i.e. CSF, clinical assessments and brain imaging, serum and plasma, and RNA. This helps identifying important biomarkers to be used for diagnosis and prognosis of PD when measurements from limited categories are available. We next use all the features listed in Table 1 to identify the most overall important PD biomarkers. We obtained the weight assigned to each feature from the learned multi-task regression model. Figure 5 shows the features with highest weight, along with their weights relative to the top-ranked feature in each group. We have the following observations regarding the feature contributions for prediction of PD progression: The first plot is obtained by including only CSF features and shows that α-Syn is the most important CSF biomarker. The potential use of α-Syn as a candidate biomarker for differentiating PD patients from healthy subjects is shown in [16]. Moreover, the evidence of a link between α-Syn at baseline and motor symptoms and cognitive speed over 2 years in PD is recently reported in [8]. However, validation of CSF α-Syn as a biomarker for PD is still an ongoing area of research. Our results validate that α-Syn is the most predictive biomaker of PD among CSF measurements. Moreover, T-tau, T-tau/$A\beta$1-42, P-tau181 and P-tau181/$A\beta$1-42 are characterized in our study as the next most important CSF markers. The study in [25] specifies the same set of CSF biomarkers as important features for differentiating between PD patients and healthy controls. We discover that the same set of CSF biomarkers are also predictive of PD status for up to 4.5 years but with different weights.

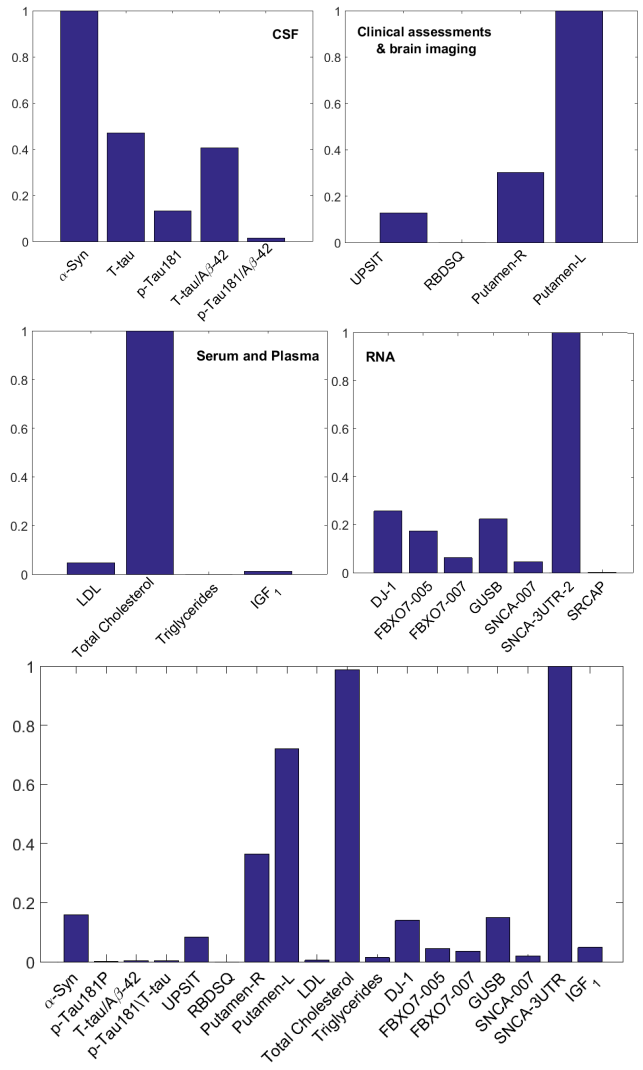

Figure 5: Feature contribution results with CSF, clinical assessments and brain imaging, serum and plasma, RNA and all the features as inputs.

Our feature contribution results on the second plot are obtained by using clinical assessments and brain imaging features and illustrates that SBR calculated on left and right putamen are the most important biomarkers in this group for prediction of PD and UPSIT assessment is the next significant feature. These results are also consistent with the feature contribution in differentiating between PD and control patients at baseline in [25].

The third plot in Figure 5 shows that total cholesterol is the biomarker with strongest contribution in prediction of PD among serum and plasma biomarkers and we identified the direction to be negative. Interestingly, a recent study published in Movement Disorders journal examined the relation of plasma lipids to Parkinson's disease in a 25 year long study with 15,792 participants and found that higher total cholesterol may be associated with lower

risk of PD [12]. However, total cholesterol is still not a validated biomarker of PD. In this study, we discovered and validated that total cholesterol is an important biomarker negatively related to progression of PD. Moreover, we identified LDL as next most important biomarker in the group of plasma and serum. LDL levels are suggested to be associated inversely with PD in [11], but the cause and effect has not been shown in that study. We validate in this study that LDL is a significant plasma biomarker of PD.

As illustrated in the fourth plot in Figure 5, we discover SCNA-3UTR (3′ untranslated region of SNCA gene) as the most promising RNA biomarker for prediction of PD. Interestingly, this biomarker has been linked to PD in a few recent genome-wide association studies [27, 30]. In fact, the SNCA gene provides instructions for making α-syn protein. We validate the significance of SCNA-3UTR α-syn in this study using a multi-task learning framework. The last plot in Figure 5 is obtained by including all the features in the model and shows that besides the important biomarkers discussed above, other RNA biomarkers including GUSB, DJ-1, FBXO7 and the serum biomarker IGF1 are also predictive of PD progression with weaker effects.

6 CONCLUSION

We studied the prediction of Parkinson's disease progression measured by unified PD rating scale (UPDRS) using baseline measurements of biologic specimen, clinical assessments and brain imaging. We particularly use a multi-task learning model where the prediction of disease status at each future visit is viewed as a task. We use a graph regularization term in the multi task regression to take into account the relatedness of PD severity at different consecutive visits and to guarantee relatively small variations between two successive visits. We performed extensive experiments on the PPMI database using different combinations of features. We also compared the results with single learner methods that do not benefit form the tasks relationships by considering different tasks independently. The results show that multi-task learning approaches outperform single learner methods in predicting the progression of PD. We also discovered a combination of RNA markers, plasma, brain imaging, CSF measurements and non-motor assessments as important biomarkers for prediction of PD progression. Specifically, SCNA-3UTR, total cholesterol, SBR in left and right putamen, α-Syn, GUSB, DJ-1 and UPSIT are identified as significant diagnostic biomarkers. We found evidence that some recent medical, genetic and neurological studies have pointed to some of our identified biomarkers such as total cholesterol, SNCA-3UTR and α-Syn. Our study validates those biomarkers and discovers new biomearkers that have not yet been reported.

The future work will focus on dealing with missing features since the current method requires a complete input matrix and observations with missing features need to be removed from the study. This is not a practical solution since it reduces the sample size and makes the framework unable to predict the future disease status for a patient with incomplete measurements.

REFERENCES

[1] Bart Bakker and Tom Heskes. 2003. Task clustering and gating for bayesian multitask learning. *Journal of Machine Learning Research* 4, May (2003), 83–99.

[2] J Booij, G Tissingh, GJ Boer, JD Speelman, JC Stoof, AG Janssen, E Ch Wolters, and EA Van Royen. 1997. [123I] FP-CIT SPECT shows a pronounced decline of striatal dopamine transporter labelling in early and advanced Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry* 62, 2 (1997), 133–140.

[3] Rich Caruana. 1997. Multitask Learning. *Mach. Learn.* 28, 1 (July 1997), 41–75. DOI: http://dx.doi.org/10.1023/A:1007379606734

[4] Rich Caruana. 1998. Multitask learning. In *Learning to learn*. Springer, 95–133.

[5] Richard L Doty, Paul Shaman, and Michael Dann. 1984. Development of the University of Pennsylvania Smell Identification Test: a standardized microencapsulated test of olfactory function. *Physiology & behavior* 32, 3 (1984), 489–502.

[6] Theodoros Evgeniou, Charles A Micchelli, and Massimiliano Pontil. 2005. Learning multiple tasks with kernel methods. *Journal of Machine Learning Research* 6, Apr (2005), 615–637.

[7] Theodoros Evgeniou and Massimiliano Pontil. 2004. Regularized multi–task learning. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 109–117.

[8] Sara Hall, Yulia Surova, Annika Öhrfelt, Henrik Zetterberg, Daniel Lindqvist, and Oskar Hansson. 2015. CSF biomarkers and clinical progression of Parkinson disease. *Neurology* 84, 1 (2015), 57–63.

[9] Tom Heskes. 2000. Empirical Bayes for learning to learn. In *ICML*. 367–374.

[10] Thomas J Hirschauer, Hojjat Adeli, and John A Buford. 2015. Computer-aided diagnosis of Parkinson's disease using enhanced probabilistic neural network. *Journal of medical systems* 39, 11 (2015), 1–12.

[11] Xuemei Huang, Robert D Abbott, Helen Petrovitch, Richard B Mailman, and G Webster Ross. 2008. Low LDL cholesterol and increased risk of Parkinson's disease: Prospective results from Honolulu-Asia Aging Study. *Movement Disorders* 23, 7 (2008), 1013–1018.

[12] Xuemei Huang, Alvaro Alonso, Xuguang Guo, David M Umbach, Maya L Lichtenstein, Christie M Ballantyne, Richard B Mailman, Thomas H Mosley, and Honglei Chen. 2015. Statins, plasma cholesterol, and risk of Parkinson's disease: a prospective study. *Movement Disorders* 30, 4 (2015), 552–559.

[13] Alex Iranzo, José Luis Molinuevo, Joan Santamaría, Mónica Serradell, María José Martí, Francesc Valldeoriola, and Eduard Tolosa. 2006. Rapid-eye-movement sleep behaviour disorder as an early marker for a neurodegenerative disorder: a descriptive study. *The Lancet Neurology* 5, 7 (2006), 572–577.

[14] Ju-Hee Kang, David J Irwin, Alice S Chen-Plotkin, Andrew Siderowf, Chelsea Caspell, Christopher S Coffey, Teresa Waligórska, Peggy Taylor, Sarah Pan, Mark Frasier, and others. 2013. 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 neurology* 70, 10 (2013), 1277–1287.

[15] Ju-Hee Kang, Brit Mollenhauer, Christopher S Coffey, Jon B Toledo, Daniel Weintraub, Douglas R Galasko, David J Irwin, Vivianna Van Deerlin, Alice S Chen-Plotkin, Chelsea Caspell-Garcia, and others. 2016. CSF biomarkers associated with disease heterogeneity in early ParkinsonâĂŹs disease: the Parkinson'âĂŹs Progression Markers Initiative study. *Acta neuropathologica* 131, 6 (2016), 935–949.

[16] Nour K Majbour, Nishant N Vaikath, Karin D van Dijk, Mustafa T Ardah, Shiji Varghese, Louise B Vesterager, Liliana P Montezinho, Stephen Poole, Bared Safieh-Garabedian, Takahiko Tokuda, and others. 2016. Oligomeric and phosphorylated alpha-synuclein as potential CSF biomarkers for Parkinson's disease. *Molecular neurodegeneration* 11, 1 (2016), 7.

[17] Kenneth Marek, Danna Jennings, Shirley Lasch, Andrew Siderowf, Caroline Tanner, Tanya Simuni, Chris Coffey, Karl Kieburtz, Emily Flagg, Sohini Chowdhury, and others. 2011. The parkinson progression marker initiative (PPMI). *Progress in neurobiology* 95, 4 (2011), 629–635.

[18] Diane B Miller and James P O'Callaghan. 2015. Biomarkers of Parkinson's disease: present and future. *Metabolism* 64, 3 (2015), S40–S46.

[19] Ziad S Nasreddine, Natalie A Phillips, Valérie Bédirian, Simon Charbonneau, Victor Whitehead, Isabelle Collin, Jeffrey L Cummings, and Howard Chertkow. 2005. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society* 53, 4 (2005), 695–699.

[20] Guillaume Obozinski, Ben Taskar, and Michael I Jordan. 2010. Joint covariate selection and joint subspace selection for multiple classification problems. *Statistics and Computing* 20, 2 (2010), 231–252.

[21] Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease and others. 2003. The Unified Parkinson's Disease Rating Scale (UPDRS): status and recommendations. *Movement disorders: official journal of the Movement Disorder Society* 18, 7 (2003), 738.

[22] Shibin Parameswaran and Kilian Q Weinberger. 2010. Large margin multi-task metric learning. In *Advances in neural information processing systems*. 1867–1875.

[23] Mirthe M Ponsen, Diederick Stoffers, Jan Booij, Berthe LF van Eck-Smit, Erik Ch Wolters, and Henk W Berendse. 2004. Idiopathic hyposmia as a preclinical sign of Parkinson's disease. *Annals of neurology* 56, 2 (2004), 173–181.

[24] RB Postuma, JF Gagnon, M Vendette, ML Fantini, J Massicotte-Marquez, and J Montplaisir. 2009. Quantifying the risk of neurodegenerative disease in idiopathic REM sleep behavior disorder. *Neurology* 72, 15 (2009), 1296–1300.

[25] R Prashanth, Sumantra Dutta Roy, Pravat K Mandal, and Shantanu Ghosh. 2016. High-Accuracy Detection of Early Parkinson's Disease through Multimodal Features and Machine Learning. *International journal of medical informatics* 90 (2016), 13–21.

[26] Novi Quadrianto, James Petterson, Tiberio S Caetano, Alex J Smola, and Svn Vishwanathan. 2010. Multitask learning without label correspondences. In *Advances in Neural Information Processing Systems*. 1957–1965.

[27] Wataru Satake, Yuko Nakabayashi, Ikuko Mizuta, Yushi Hirota, Chiyomi Ito, Michiaki Kubo, Takahisa Kawaguchi, Tatsuhiko Tsunoda, Masahiko Watanabe, Atsushi Takeda, and others. 2009. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nature genetics* 41, 12 (2009), 1303–1307.

[28] Anette Schrag, Uzma Faisal Siddiqui, Zacharias Anastasiou, Daniel Weintraub, and Jonathan M Schott. 2017. Clinical variables and biomarkers in prediction of cognitive impairment in patients with newly diagnosed Parkinson's disease: a cohort study. *The Lancet Neurology* 16, 1 (2017), 66–75.

[29] Leslie M Shaw, Hugo Vanderstichele, Malgorzata Knapik-Czajka, Christopher M Clark, Paul S Aisen, Ronald C Petersen, Kaj Blennow, Holly Soares, Adam Simon, Piotr Lewczuk, and others. 2009. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Annals of neurology* 65, 4 (2009), 403–413.

[30] Javier Simon-Sanchez, Claudia Schulte, Jose M Bras, Manu Sharma, J Raphael Gibbs, Daniela Berg, Coro Paisan-Ruiz, Peter Lichtner, Sonja W Scholz, Dena G Hernandez, and others. 2009. Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nature genetics* 41, 12 (2009), 1308–1312.

[31] Karin Stiasny-Kolster, Geert Mayer, Sylvia Schäfer, Jens Carsten Möller, Monika Heinzel-Gutenbrunner, and Wolfgang H Oertel. 2007. The REM sleep behavior disorder screening questionnaire new diagnostic instrument. *Movement Disorders* 22, 16 (2007), 2386–2393.

[32] Robert Tibshirani, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. 2005. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67, 1 (2005), 91–108.

[33] Martine Visser, Johan Marinus, Anne M Stiggelbout, and Jacobus J Van Hilten. 2004. Assessment of autonomic dysfunction in Parkinson's disease: The SCOPA-AUT. *Movement Disorders* 19, 11 (2004), 1306–1312.

[34] Christian Widmer and Gunnar Rätsch. 2012. Multitask Learning in Computational Biology.. In *ICML Unsupervised and Transfer Learning*. 207–216.

[35] Ya Xue, Xuejun Liao, Lawrence Carin, and Balaji Krishnapuram. 2007. Multi-task learning for classification with Dirichlet process priors. *Journal of Machine Learning Research* 8, Jan (2007), 35–63.

[36] Yu Zhang and Dit-Yan Yeung. 2010. Multi-Task Learning using Generalized t Process.. In *AISTATS*. 964–971.

[37] Jiayu Zhou, Jianhui Chen, and Jieping Ye. 2011. Malsar: Multi-task learning via structural regularization. *Arizona State University* (2011).

[38] Jiayu Zhou, Jun Liu, Vaibhav A Narayan, and Jieping Ye. 2012. Modeling disease progression via fused sparse group lasso. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1095–1103.

[39] Jiayu Zhou, Lei Yuan, Jun Liu, and Jieping Ye. 2011. A multi-task learning formulation for predicting disease progression. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 814–822.