Parkinson's disease diagnosis via joint learning from multiple modalities and relations

Lei, Haijun; Huang, Zhongwei; Zhou, Feng; Elazab, Ahmed; Tan, Ee-Leng; Li, Hancong; Qin, Jing; Lei, Baiying

2019

Lei, H., Huang, Z., Zhou, F., Elazab, A., Tan, E.-L., Li, H., . . . Lei, B. (2019). Parkinson's disease diagnosis via joint learning from multiple modalities and relations. *IEEE Journal of Biomedical and Health Informatics*, 23(4), 1437-1449. doi:10.1109/JBHI.2018.2868420

<https://hdl.handle.net/10356/137182>

<https://doi.org/10.1109/JBHI.2018.2868420>

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at: <https://doi.org/10.1109/JBHI.2018.2868420>.

Downloaded on 26 Apr 2023 20:34:10 SGT

Parkinson's Disease Diagnosis via Joint Learning from Multiple Modalities and Relations

Haijun Lei, Zhongwei Huang, Feng Zhou, Ahmed Elazab, Ee-Leng Tan, Hancong Li, Jing Qin, Baiying Lei*, *Senior Member, IEEE*

Abstract—Parkinson's disease (PD) is a neurodegenerative progressive disease that mainly affects the motor systems of patients. To slow this disease deterioration, early and accurate diagnosis of PD is an effective way, which alleviates mental and physical sufferings by clinical intervention. In this paper, we propose a joint regression and classification framework for PD diagnosis via magnetic resonance and diffusion tensor imaging data. Specifically, we devise a unified multi-task feature selection model to explore multiple relationships among features, samples, and clinical scores. We regress four clinical variables of depression, sleep, olfaction, cognition scores, as well as perform the classification of PD disease from the multi-modal data. The multi-task model explores the relationships at the level of clinical scores, image features, and subjects, to select the most informative and diseased-related features for diagnosis. The proposed method is evaluated on the public Parkinson's progression markers initiative (PPMI) dataset. The extensive experimental results show that the multi-task framework can effectively boost the performance of regression and classification and outperforms other state-of-the-art methods. The computerized predictions of clinical scores and label for PD diagnosis may offer quantitative reference for decision support as well.

Index Terms—Parkinson's disease diagnosis; Joint learning; Multiple relations; Multi-modal; Classification; Prediction.

I. INTRODUCTION

Parkinson's disease (PD) is a neurodegenerative and irreversible disorder in the elderly people. Patients with PD in the middle and late stages endure endless mental and

physical sufferings due to the progressive occurrence of diseases. In general, PD can be mainly identified with the appearing of the four motor symptoms of tremor, rigidity, bradykinesia, and postural instability as well as the four non-motor symptoms of depression, sleep, olfaction, and cognition disorders [1-4]. These symptoms are mainly caused by the death of dopamine neurons in the substantia nigra [5]. However, there are approximately 10% of patients clinically diagnosed with early PD, which normally have dopaminergic functional imaging (scans without evidence of dopamine deficit, SWEDDs) [6, 7]. This situation has undoubtedly increased the challenges of PD diagnosis. Although some surgeries can effectively relieve these symptoms for advanced PD, they all have some side effects. It is known that unilateral subthalamotomy [8, 9] is effective to mitigate the primary motor features of PD and improve the patients' mood. However, this procedure is invasive. Also, the right subthalamotomy is at the high risk of impulsivity, while the left subthalamotomy causes further impairment of semantic verbal fluency. In fact, deep brain stimulation of the subthalamic nucleus [10] is a well-established treatment for PD. However, the spread of stimulation to surrounding structures causes the adverse effects, which will affect the functionality of these areas. For this reason, monitoring disease progression and alleviating patients' mental and physical suffering are highly desirable for early PD diagnosis.

Recently, the computer-aided PD diagnosis from neuroimaging data has played an increasingly vital role and attracted much attention [11, 12]. For example, Prashanth *et al.* [13] employed the striatal binding ratio values from single photon emission computed tomography images for automated classification and prediction. Abós *et al.* [14] used functional connectivity from resting-state functional magnetic resonance imaging (fMRI) as feature for classification. Liu *et al.* [15] proposed a novel feature selection method via iterative canonical correlation analysis, which fuses features of different types into a common space to find the optimal feature subset from two sets of features. Since multiple modalities can provide more comprehensive information, they have been widely used in the recent neurodegenerative diseases studies [16]. For instance, Adeli *et al.* [17] introduced a kernel-based feature selection scheme to best benefit the classification scheme in the kernel space instead of the original input feature space. Recently, Adeli *et al.* [18] proposed a semi-supervised robust discriminative classification method using both labeled training and unlabeled testing data to detect the sample-outliers and feature-noises, simultaneously. Inspired by this, we explore multiple modalities (e.g., gray matter (GM) and cerebrospinal

This work was supported partly by National Natural Science Foundation of China (Nos. 61871274, 61801305 and 61571304), Guangdong Pre-national project (No. 2014GKXM054), National Natural Science Foundation of Guangdong Province (Nos. 2017A030313377 and 2016A030313047), the Integration Project of Production Teaching and Research by Guangdong Province and Ministry of Education (No. 2012B091100495), Shenzhen Peacock Plan (No. KQTD2016053112051497), and Shenzhen Key Basic Research Project (Nos. JCYJ20120613113419607, JCYJ20150930105133185, and JCYJ20170302153337765). Asterisk indicates corresponding author.

H. Lei, Z. Huang and H. Li are with the Key Laboratory of Service Computing and Applications, Guangdong Province Key Laboratory of Popular High Performance Computers, College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China, 518060.

A. Elazab and B. Lei are with the National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China, 518060 (e-mail: leiby@szu.edu.cn).

F. Zhou is with Industrial and Manufacturing Systems Engineering, University of Michigan at Dearborn, Dearborn, MI, USA, 48128.

E-L Tan is with School of Electronic and Electric Engineering, Nanyang Technological University, Singapore, 639798.

J. Qin is with School of Nursing, Hong Kong Polytechnic University, Hung Hom, Hong Kong.

fluid (CSF) from magnetic resonance imaging (MRI), and mean diffusion (MD) coefficient of diffusion tensor imaging (DTI) [19]) to discriminate PD, SWEDD, and NC. To further boost the diagnosis performance, we compute the weighted summation (voting) of the prediction results of the three features (i.e., GM, CSF, and MD (GCM)) as the final prediction result. The main reason is that, each feature contributes uniquely to the final prediction, and the majority voting is able to enhance the performance by exploring the feature complementarity.

However, for the multi-modal neuroimaging data, the feature dimension is relatively high, while the number of the subjects is quite limited, which easily leads to overfitting and makes it difficult to generate a robust model [20]. To address it, feature selection is an effective way by discovering the disease-related characteristics. Inspired by promising performance in [21], a multi-relational regularization scheme with the improved loss function is devised to jointly obtain disease-related features. Specifically, we incorporate the relationships among features, subjects and clinical scores in a joint learning model for performance boosting. By establishing this model, it is of great importance to accurately determine the treatment planning rather than the general treatment in the clinical diagnosis. However, there are few approaches proposed for this task. Since clinical scores prediction can help doctors to stage the disease with the proper treatment planning, they should be jointly predicted in our framework.

Despite some good achievements have been made in literature, our work is novel and different from others in the following aspects. Firstly, existing methods mainly utilize single modality features for PD diagnosis [13-15]. Although some methods already used multi-modal features [21, 22], these methods only consider linear connections among features without considering feature complementarity. Therefore, the feature representation is quite weak. Since each feature contributes uniquely to the final prediction and the majority voting is able to enhance the performance by exploring the feature complementarity, we compute the weighted summation (voting) of the prediction results of multi-modal features (i.e., GM, CSF, and MD) as the final prediction result. Secondly, most existing studies mainly concentrate on the classification task in PD diagnosis [15, 17, 18]. However, apart from showing some motor symptoms, patients with PD also exhibit some non-motor symptoms (depression, sleep, olfaction and cognitive impairment) in clinical diagnosis. In other words, the brain mechanism can simultaneously affect clinical scores and prediction labels. Therefore, in this paper, we devise a unified multi-task feature selection method for simultaneous classification and prediction. Clinical scores predictions not only provide a good guide to classification, but also help doctors to classify disease stages with the proper treatment planning. Thirdly, most existing studies mainly focus on either separate response variables regularization or data regularization. For instance, in [22], authors simultaneously performed classification and clinical scores prediction based on an improved loss function that considers the relations among rows

and the information among columns in response variables. However, it failed to consider data regularization. In [21], authors utilized multi-modal data to improve the classification performance of Alzheimer's disease via discriminative learning and relational regularization (feature-feature and subject-subject similarities) and did not consider response variables regularization. Zhu *et al.* [23] simultaneously considered response variables regularization and data regularization. However, the response variables similarity constraint in the method may not accurately reflect the structural information of the responses variable since the similarity between label and clinical score is weak. Hence, in the paper, we devise a unified multi-task feature selection model to explore multiple relations among features, samples, and response variables (i.e., feature-feature similarity, subject-subject similarity, and similarities of column and row differences in response variables) for joint classification and scores prediction. Specifically, we adopt the values of the geriatric depression scale (GDS) [24] as the depression scores, the values of the Epworth sleepiness scale (ESS) [25] as the sleep scores, the values of the university of Pennsylvania smell identification test (UPSIT) [26] as the olfaction scores, and the values of the Montreal cognitive assessment (MoCA) [27] as the cognition scores based on the specification from the Parkinson's progression markers initiative (PPMI) dataset. We simultaneously perform classification and clinical scores prediction using multi-modal features and multiple relationships. We utilize the depression, sleep, olfaction, cognition scores, and a clinical label to jointly select the most discriminative features. The extensive experiments have been performed to validate the proposed method. The promising diagnosis performance achieved on the public PPMI dataset shows that our proposed method is feasible for practical application of PD diagnosis.

II. DATASET AND PREPROCESSING

A. Data Acquisition

In this paper, we explore the public PPMI database [28] for data acquisition since it is the international database to identify PD progression biomarker. Both MRI and DTI data are utilized, which are collected from the Siemens MAGNETOM trio 3.0T MRI scanner. For MRI images, the data acquisition parameters are set as follows: acquisition plane=SAGITTAL, pulse sequence=GR/IR, field strength=3 tesla, slice thickness=1 mm, flip angle=9 degree, TE=2.98 ms, and TR=2300 ms. We select the DTI images using the following parameters: pulse sequence=EP, gradient directions =64, field strength=3 tesla, slice thickness=2 mm, flip angle=90 degree, TE=88 ms, and TR=600-1000 ms.

B. Subjects

A total of 238 subjects including 62 NC subjects, 142 PD subjects, and 34 SWEDD subjects are used in this study. The depression, sleep, olfaction, and cognition scores are evaluated by GDS, ESS, UPSIT, and MoCA, respectively. Specifically, the depression scores are collected from a yes/no survey's answer from GDS. The depression ranges from: 0 to 4, 5 to 7, 8

to 11, and 12 to 15 are considered as normal, slight, medium and serious, respectively. The sleep scores are evaluated by the sum of weighted responses from several questions from ESS. The sleep ranges from 0 to 9 is normal while 10 to 24 is considered sleepy. For olfaction scores, the range of raw olfaction score of UPSIT is between 0 and 40. Lower olfaction scores indicate that the subject has lost much of their smell sense. For cognition scores, one point is added to the score for a subject who has 12 years or below of formal education. Each subject can be scored a maximum of 30 points. The simple sum of both individual question scores and total score is available. Table 1 summarizes the clinical information of the subjects.

Meanwhile, to evaluate the statistical significance of the clinical information, we implement the paired t -tests (95% significance level) on age, gender, weight, depression score, sleep score, olfaction score, and cognition score, respectively. For NC and PD, since the number of PD is larger than NC, we randomly take the subjects equal to NC from the PD and then calculate the p -values. For fair evaluation, we repeat the above process ten times and the resulting p -values indicate that there are statistical differences in other clinical information except for gender (with p -value > 0.05). Similarly, for NC and SWEDD, the resulting p -values also show that there are statistical differences in other clinical information except for gender (with p -value > 0.05). For PD and SWEDD, the resulting p -values indicate that there are statistical differences in other clinical information except for gender and weight (with p -values > 0.05).

TABLE I
CLINICAL DETAILS OF ALL SUBJECTS (MEAN $\pm$ STANDARD DEVIATION).

	NC	PD	SWEDD
Number	62	142	34
Female/male	23/39	51/91	14/20
Age	61.0 $\pm$ 10.7	61.5 $\pm$ 9.4	61.7 $\pm$ 9.8
Weight (Kg)	76.9 $\pm$ 15.4	82.6 $\pm$ 16.6	83.1 $\pm$ 12.5
Depression scores	5.1 $\pm$ 1.2	5.3 $\pm$ 1.5	5.6 $\pm$ 1.3
Sleep scores	6.5 $\pm$ 3.9	5.9 $\pm$ 3.4	8.7 $\pm$ 4.2
Olfaction scores	33.4 $\pm$ 4.7	22.5 $\pm$ 8.5	29.9 $\pm$ 8.4
Cognition scores	28.2 $\pm$ 1.1	27.5 $\pm$ 2.1	26.9 $\pm$ 2.9

C. Data Preprocessing

As for data preprocessing, we first perform anterior commissure-posterior commissure reorientation in all MRI images using center of mass algorithm. The statistical parametric mapping (SPM8) (<http://www.fil.ion.ucl.ac.uk/spm>) is applied to correct the geometric distortion and head movement. Then, we implement skull-stripping using graph-cut method [29]. Afterwards, we register the MRI images with international consortium for brain mapping template, which serves a reference providing coordinates and the associated anatomical labels and divides the associated regions into GM and CSF. Meanwhile, all images are resampled to the isotropic resolution of 1.5mm to make the resolution invariant. In addition, we exploit 60-mm full width at half maximum Gaussian kernel to spatially smooth the surface of MRI images. Finally, we register (automated anatomical labeling) AAL atlas [30] with GM and CSF, respectively, and then separately extract 116 features from GM and CSF using the registered AAL templates.

For DTI data, each sample has 65 images (DICOM format), including a b0 image with no diffusion gradient active and 64 images with various gradient directions. First, FSL tool [31] corrects the b0 distortion using b0 image. Second, we input the corrected b0 image into *bet* function of FSL to generate a mask. Third, we use *dcm2nii* tool to convert the original 65 images into a 4D data and simultaneously generate a b-value file containing a scalar value for each gradient and a b-vector file indicating each gradient direction. Fourth, we input 4D image to *eddy_correct* function of FSL to conduct eddy current correction on the images. Fifth, we input mask image, corrected 4D image, b-value and b-vector files to *dtifit* function of FSL to generate MD data. Finally, we register AAL atlas with MD and then extract 116 features from MD using the registered AAL templates.

III. METHODOLOGY

A. System Overview

The overall procedures of the proposed method are presented in Fig.1. We will describe it in the following subsections in detail.

B. Notations

In this study, the uppercase boldface letters (e.g., $\mathbf{X}$) denote matrices, and lowercase boldface letters denote vectors. For a matrix $\mathbf{X} = [x_{kl}]$, where $\mathbf{x}^k$ and $\mathbf{x}_l$ denote its k -th row and l -th column, respectively, the trace of matrix $\mathbf{X}$ is denoted as $\text{tr}(\mathbf{X})$. The norm of vector is defined as $\|\mathbf{X}\|$. We denote the Frobenius norm and $\ell_{2,1}$ -norm of a matrix $\mathbf{X}$ as $\|\mathbf{X}\|_F = \sqrt{\sum_k \|\mathbf{x}^k\|_2^2}$, and $\|\mathbf{X}\|_{2,1} = \sum_k \|\mathbf{x}^k\|_2$, respectively.

C. Proposed Method

Let $\mathbf{X} \in \mathbf{R}^{d \times n}$ and $\mathbf{Y} \in \mathbf{R}^{s \times n}$ denote the training data and the target response matrix of n subjects, d features, and s response scores, respectively. The response matrix $\mathbf{Y}$ is obtained by concatenating clinical scores (depression, sleep, olfaction, and MoCA scores) and class label of different samples (NC, PD, and SWEDD). In general, a least square regression model for joint regression and classification is defined as

$$\mathcal{F}(\mathbf{W}) = \|\mathbf{Y} - \hat{\mathbf{Y}}\|_F^2 = \|\mathbf{Y} - \mathbf{W}^T \mathbf{X}\|_F^2, \quad (1)$$

where $\mathbf{W} \in \mathbf{R}^{d \times s}$ is the weight coefficient matrix and each column of $\mathbf{W}$ contains different weight coefficients of each feature and $\hat{\mathbf{Y}}$ is the predicted response matrix. For the neuroimaging dataset, it has small subjects and high-dimensional features, which causes the overfitting issue. To handle it, we use the regularization terms to enhance the generalization ability. The model with the regularization term is expressed as

$$\min_{\mathbf{W}} \mathcal{F}(\mathbf{W}) + \mathcal{R}(\mathbf{W}), \quad (2)$$

where $\mathcal{R}(\mathbf{W})$ denotes the regularization term.

For the target response matrix $\mathbf{Y}$ and the predicted response matrix $\hat{\mathbf{Y}}$, there is an inherent information to measure the similarities, such as the relationships among any pair of columns and/or any pair of rows in both matrices. Specifically, the column-wise relation means that the difference of any pair of columns in $\mathbf{Y}$ is similar to the corresponding columns in $\hat{\mathbf{Y}}$, namely, the similarity between $(\mathbf{y}_i - \mathbf{y}_j)$ and $(\hat{\mathbf{y}}_i - \hat{\mathbf{y}}_j)$.

Fig.1. Illustration of the proposed method (Note that; large rectangles with different colors mean diverse features and response variables, respectively. Rectangle surrounded by dotted line indicates additional clinical features, star denotes column-wise relation in response variables, triangle represents row-wise relation in response variables, circle means feature-feature relation, and rhombus denotes subject-subject relation.).

The row-wise relation means that the difference of any pair of rows in $\mathbf{Y}$ is similar to the corresponding rows in $\hat{\mathbf{Y}}$, i.e., the similarity between $(\mathbf{y}^i - \mathbf{y}^j)$ and $(\hat{\mathbf{y}}^i - \hat{\mathbf{y}}^j)$. Meanwhile, we also can easily notice that the two regularization terms are reasonable from the perspective of graph theory. For example, we treat n column vectors of $\mathbf{Y}$ as n points and the Euclidean distance between any pair of points is used as the weight of the edge. By this way, we generate the graph $\mathbf{A}$ of column vectors of $\mathbf{Y}$. Similarly, we also construct the graph $\mathbf{B}$ of the predicted response matrix $\hat{\mathbf{Y}}$. According to our intuitive understanding, we need $\mathbf{A}$ and $\mathbf{B}$ to be as similar as possible, that is, $(\mathbf{y}_i - \mathbf{y}_j)$ and $(\hat{\mathbf{y}}_i - \hat{\mathbf{y}}_j)$ are similar. In the same way, we can reasonably explain the second regularization term by constructing the graph in the above way. The two kinds of relations among response variables are defined as the two regularization terms

$$\mathcal{R}_1(\mathbf{W}) = \lambda_1 \sum_{i,j}^n \|(\mathbf{y}_i - \mathbf{y}_j) - (\hat{\mathbf{y}}_i - \hat{\mathbf{y}}_j)\|_2^2 \\ = \lambda_1 \sum_{i,j}^n \|(\mathbf{y}_i - \mathbf{y}_j) - (\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j)\|_2^2, \quad (3)$$

$$\mathcal{R}_2(\mathbf{W}) = \lambda_2 \sum_{i,j}^s \|(\mathbf{y}^i - \mathbf{y}^j) - (\hat{\mathbf{y}}^i - \hat{\mathbf{y}}^j)\|_2^2 \\ = \lambda_2 \sum_{i,j}^s \|(\mathbf{y}^i - \mathbf{y}^j) - (\mathbf{w}_i^T \mathbf{X} - \mathbf{w}_j^T \mathbf{X})\|_2^2, \quad (4)$$

where λ_1 and λ_2 are the controlling parameters of the regularization terms. The lower value of the regularization term, the more similar they are.

In fact, there are relations among features extracted from ROIs. If two features are similar, the corresponding weight coefficients are similar. If some features, e.g., $\mathbf{x}^i$ and $\mathbf{x}^j$, are related to each other, their corresponding weight coefficients (i.e., $\mathbf{w}^i$ and $\mathbf{w}^j$) would be similar since the k -th feature $\mathbf{x}^i$ in $\mathbf{X}$ corresponds to the k -th row $\mathbf{w}^k$ in $\mathbf{W}$ in our regression framework. To utilize this relationship, we devise a new regularization term using the relationships among features, which is denoted as

$$\mathcal{R}_3(\mathbf{W}) = \lambda_3 \sum_{i,j}^d h_{ij} \|\mathbf{w}^i - \mathbf{w}^j\|_2^2, \quad (5)$$

where λ_3 is the controlling parameter and h_{ij} represents an element in the feature similarity matrix $\mathbf{H} = [h_{ij}] \in \mathbb{R}^{d \times d}$, which quantizes the relation among features.

Also, if some samples $\mathbf{x}_i$ and $\mathbf{x}_j$ are related to each other, the corresponding predicted response variables $\hat{\mathbf{y}}_i$ and $\hat{\mathbf{y}}_j$ shall be similar since $\mathbf{W}^T \mathbf{x}_i$ and $\mathbf{W}^T \mathbf{x}_j$ corresponds to $\hat{\mathbf{y}}_i$ and $\hat{\mathbf{y}}_j$, respectively. We incorporate this relation as an additional regularization term as below

$$\mathcal{R}_4(\mathbf{W}) = \lambda_4 \sum_{i,j}^n v_{ij} \|\hat{\mathbf{y}}_i - \hat{\mathbf{y}}_j\|_2^2 = \lambda_4 \sum_{i,j}^n v_{ij} \|\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j\|_2^2, \quad (6)$$

where λ_4 is the regularization parameter, v_{ij} represents an element in the subject similarity matrix $\mathbf{V} = [v_{ij}] \in \mathbb{R}^{n \times n}$, which quantizes the relation among subjects.

To measure the similarity between vectors $\mathbf{a}$ and $\mathbf{b}$, we exploit a radial basis function kernel, which is given by

$$g(\mathbf{a}, \mathbf{b}) = \exp\left(-\frac{\|\mathbf{a} - \mathbf{b}\|_2^2}{2\sigma^2}\right), \quad (7)$$

where σ denotes the kernel width. We build a data adjacency graph and regard each vector as a node using k -nearest neighbors along with the kernel function defined in Eq. (7) to compute the edge weights. For the similarity matrix $\mathbf{V}$, if a sample $\mathbf{x}_i$ is chosen as one of the k -nearest neighbors of a sample $\mathbf{x}_j$, then the similarity v_{ij} between the two nodes is set to the value of $g(\mathbf{x}_i, \mathbf{x}_j)$, otherwise, $v_{ij} = 0$. Similarly, h_{kl} is computed as v_{kl} as described above.

Since the brain mechanisms can simultaneously affect the clinical scores and class labels, we use the same features for class label identification and clinical scores prediction formulated by a $\ell_{2,1}$ -norm regularization term on $\mathbf{W}$. Finally, our loss function is formulated as

$$\min_{\mathbf{W}} \mathcal{F}(\mathbf{W}) + \mathcal{R}_1(\mathbf{W}) + \mathcal{R}_2(\mathbf{W}) + \mathcal{R}_3(\mathbf{W}) + \mathcal{R}_4(\mathbf{W}) + \lambda_5 \|\mathbf{W}\|_{2,1}, \quad (8)$$

where λ_5 denotes the weighting parameter.

D. Optimization

For the optimization of parameters $\mathbf{W}$, it is necessary to transform the regularization terms to the trace forms using Laplacian matrices. Let $\mathbf{M}^C$, $\mathbf{M}^D$, $\mathbf{M}^H$, and $\mathbf{M}^V$ be diagonal matrices (i.e., degree matrices), where the diagonal values are computed by the row-wise sum of the similarity weight matrices of $\mathbf{C}$, $\mathbf{D}$, $\mathbf{H}$, and $\mathbf{V}$, respectively (i.e., $m_{ii}^C = \sum_{j=1}^n 1$, $m_{ii}^D = \sum_{j=1}^s 1$, $m_{ii}^H = \sum_{j=1}^d h_{ij}$, and $m_{ii}^V = \sum_{j=1}^n v_{ij}$). The regularization terms can be equivalently transformed as

$$\mathcal{R}_1(\mathbf{W}) = \lambda_1 \sum_{i,j}^n \|(\mathbf{y}_i - \mathbf{y}_j) - (\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j)\|_2^2 \\ = \lambda_1 \sum_{i,j}^n \|(\mathbf{y}_i - \mathbf{W}^T \mathbf{x}_i) - (\mathbf{y}_j - \mathbf{W}^T \mathbf{x}_j)\|_2^2 \\ = \lambda_1 \text{tr}(\mathbf{W}^T \mathbf{X} \mathbf{L}_C \mathbf{X}^T \mathbf{W} - 2 \mathbf{Y} \mathbf{L}_C \mathbf{X}^T \mathbf{W}), \quad (9)$$

$$\mathcal{R}_2(\mathbf{W}) = \lambda_2 \sum_{i,j}^s \|(\mathbf{y}^i - \mathbf{y}^j) - (\mathbf{w}_i^T \mathbf{X} - \mathbf{w}_j^T \mathbf{X})\|_2^2 \\ = \lambda_2 \sum_{i,j}^s \|(\mathbf{y}^i - \mathbf{w}_i^T \mathbf{X}) - (\mathbf{y}^j - \mathbf{w}_j^T \mathbf{X})\|_2^2 \\ = \lambda_2 \text{tr}(\mathbf{X}^T \mathbf{W} \mathbf{L}_R \mathbf{W}^T \mathbf{X} - 2 \mathbf{X}^T \mathbf{W} \mathbf{L}_R \mathbf{Y}), \quad (10)$$

$$\mathcal{R}_3(\mathbf{W}) = \lambda_3 \text{tr}(\mathbf{W} \mathbf{L}_H \mathbf{W}^T), \quad (11)$$

$$\mathcal{R}_4(\mathbf{W}) = \lambda_4 \text{tr}(\mathbf{W}^T \mathbf{X} \mathbf{L}_V \mathbf{X}^T \mathbf{W}), \quad (12)$$

where $\mathbf{L}_C = \mathbf{M}^C - \mathbf{E}_n$, $\mathbf{L}_R = \mathbf{M}^R - \mathbf{E}_s$, $\mathbf{L}_H = \mathbf{M}^H - \mathbf{M}$, and $\mathbf{L}_V = \mathbf{M}^V - \mathbf{V}$ are Laplacian matrices. $\mathbf{E}_n$ and $\mathbf{E}_s$ denote

identity matrices of size n and s , respectively. Finally, our objective function in Eq. (8) can be rewritten as follows:

$$\begin{aligned} \min_{\mathbf{W}} \mathcal{F}(\mathbf{W}) + \lambda_1 \text{tr}(\mathbf{W}^T \mathbf{X} \mathbf{L}_C \mathbf{X}^T \mathbf{W} - 2 \mathbf{Y} \mathbf{L}_C \mathbf{X}^T \mathbf{W}) \\ + \lambda_2 \text{tr}(\mathbf{X}^T \mathbf{W} \mathbf{L}_R \mathbf{W}^T \mathbf{X} - 2 \mathbf{X}^T \mathbf{W} \mathbf{L}_R \mathbf{Y}) + \lambda_3 \text{tr}(\mathbf{W} \mathbf{L}_H \mathbf{W}^T) \\ + \lambda_4 \text{tr}(\mathbf{W}^T \mathbf{X} \mathbf{L}_V \mathbf{X}^T \mathbf{W}) + \lambda_5 \|\mathbf{W}\|_{2,1}. \end{aligned} \quad (13)$$

By setting the derivative of the loss function on $\mathbf{W}$ in Eq. (13) to zero, we obtain a Lyapunov equation

$$\mathbf{A} \mathbf{W} + \mathbf{W} \mathbf{B} = \mathbf{C}, \quad (14)$$

where $\mathbf{A} = (\mathbf{X} \mathbf{X}^T)^{-1} (\mathbf{X} \mathbf{X}^T + 2 \lambda_1 \mathbf{X} \mathbf{L}_C \mathbf{X}^T + \lambda_3 \mathbf{L}_H + \lambda_4 \mathbf{X} \mathbf{L}_V \mathbf{X}^T + \lambda_5 \mathbf{Z})$, $\mathbf{B} = 2 \lambda_2 \mathbf{L}_R$, $\mathbf{C} = (\mathbf{X} \mathbf{X}^T)^{-1} (\mathbf{X} \mathbf{Y}^T + 2 \lambda_1 \mathbf{X} \mathbf{L}_C \mathbf{Y}^T + 2 \lambda_2 \mathbf{X} \mathbf{Y}^T \mathbf{L}_R)$, and $\mathbf{Z} \in \mathbf{R}^{d \times d}$ is a diagonal matrix, where the i -th diagonal value is given as

$$z_{ij} = \frac{1}{2 \|\mathbf{w}^i\|_2}. \quad (15)$$

To solve Eq. (14), we use an iterative method to optimize it since $\mathbf{Z}$ and $\mathbf{W}$ are interdependent. First, we use an identity matrix to initialize the matrix $\mathbf{Z}(0)$. Second, we update the matrix $\mathbf{W}(1)$ with the matrix $\mathbf{Z}(0)$ and then update the matrix $\mathbf{Z}(1)$ with the matrix $\mathbf{W}(1)$, iteratively. After 50 iterations, it generally converges. So, we execute this iterative process until the iteration number t reaches 50. Finally, the matrix $\mathbf{Z}$ is updated. Since some values may appear indeterminate or infinite, we set these values to *eps*. Algorithm 1 summarizes the implementation steps of solving Eq. (14).

E. Model Training

Due to the $\ell_{2,1}$ -norm regularization term in the objective function, some rows are zero or close to zero in the matrix $\mathbf{W}$ after optimizing the solution in algorithm 1. Each column of $\mathbf{W}$ corresponds to different weighted coefficients of each feature in the sample, the greater the weight, the more relevant the feature corresponds to the specific task. In $\mathbf{W}$, there are five columns for depression, sleep, olfaction, and cognition scores prediction and classification. To select the shared features associated with these tasks, we obtain the column vector $\mathbf{p}$ by computing the square value of ℓ_2 -norm of each row in the matrix $\mathbf{W}$ and then gain the mean value c of $\mathbf{p}$. According to the empirical result, if the values in the column vector $\mathbf{p} \geq 0.2 \times c$, we will select the corresponding features of these values. The selected features are appended to train support vector machines via LIBSVM (<https://www.csie.ntu.edu.tw/~cjlin/libsvm/>), which build support vector regression (SVR) models for predicting depression, sleep, olfaction, and cognition scores and support vector classification (SVC) models to discriminate clinical labels.

Algorithm 1: Pseudo code of solving Eq. (14)

Input: $\mathbf{X} \in \mathbf{R}^{d \times n}$, $\mathbf{Y} \in \mathbf{R}^{c \times n}$, $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5$;
Output: $\mathbf{W} \in \mathbf{R}^{d \times c}$;

- 1 Initialize $t = 0$, $\mathbf{Z}(t)$, $\mathbf{W}(t)$ by an identity matrix;
- 2 **Repeat**
- 3 Update matrix $\mathbf{W}(t + 1)$ by solving Eq. (14);
- 4 Update matrix $\mathbf{Z}(t + 1)$ by solving Eq. (15);
- 5 If $\mathbf{Z}(t + 1)$ is indeterminate or infinite
 $\mathbf{Z}(t + 1) = \text{eps}$
- 6 End
- 7 $t = t + 1$;
- 7 **Until** it converges or a stop criterion is satisfied.

IV. EXPERIMENTAL RESULTS

A. Experimental Setting

In our study, we conduct three binary classification experiments: NC vs. PD, NC vs. SWEDD, and PD vs. SWEDD. In the meantime, we consider four scores prediction (depression, sleep, olfaction, and cognition scores prediction). Our experiments are conducted using 10-fold cross-validation to validate the performance of the proposed method. Specifically, we divide our dataset randomly into ten subsets, where ten percent of the dataset is used for testing and the remaining is used for training. We repeat this process ten times to have the generalization results. The final results are obtained by averaging the repeated experimental results. We tune parameters for model selection. The SVR/SVC parameters are as follows: $C \in \{2^{-10}, \dots, 2^{10}\}$ and $G \in \{10^{-5}, \dots, 10^5\}$. In NC vs. PD, we conduct the grid search on the parameter spaces of $\lambda_i \in \{10^{-5}, \dots, 10^2\}$, $i \in \{1, 2\}$, $\lambda_j \in \{10^{-8}, \dots, 10^{-5}\}$, $j \in \{3, 4\}$, $\lambda_5 = 100$ in GM and CSF data and $\lambda_i \in \{10^{-6}, \dots, 10^2\}$, $i \in \{1, 2, 4\}$, $\lambda_3 \in \{10^2, \dots, 10^8\}$, $\lambda_5 = 10^4$ in MD data. In NC vs. SWEDD, we carry out the grid search on the parameter spaces of $\lambda_i \in \{10^{-6}, \dots, 10^2\}$, $i \in \{1, 2, 4\}$, $\lambda_3 \in \{10^2, \dots, 10^8\}$, $\lambda_5 = 10^4$ in GM and CSF data, and $\lambda_i \in \{10^{-5}, \dots, 10^2\}$, $i \in \{1, 2, 3, 4\}$, $\lambda_5 = 10^3$ in MD data. In PD vs. SWEDD, we execute the grid search on the parameter spaces of $\lambda_i \in \{10^{-5}, \dots, 10^2\}$, $i \in \{1, 2, 3, 4\}$, $\lambda_5 = 10^3$ in GM, CSF, and MD data. In GCM data, we also implement the grid search on the parameter spaces of $\lambda_i \in \{10^{-5}, \dots, 10^2\}$, $i \in \{1, 2, 3, 4\}$, $\lambda_5 = 10^3$. In NC vs. PD, the voting weights corresponding to GM, CSF and MD data for the classification and the regression predictions are 0.09, 0.01, and 0.9, respectively. In NC vs. SWEDD, the voting weights for the classification are 0.1, 0.8, and 0.1, respectively, and the voting weights for the regression predictions are 0.09, 0.9, and 0.01, respectively. In PD vs. SWEDD, the voting weights for the classification are 0.1, 0.1, and 0.8, respectively, and the voting weights for the regression predictions are 0.01, 0.09, and 0.9, respectively. Meanwhile, we empirically set $k=3$ and $\sigma=1$ to compute the similarity among features and subjects. Also, we append the clinical features (age and sex) [32-35] to jointly predict four clinical scores.

B. Comparison Methods

We compare the proposed method with the following state-of-the-art methods.

Least absolute shrinkage and selection operator method (Lasso) [36] is integrated into MATLAB statistical toolbox as a regularization technique useful for feature selection to avoid over-fitting of training data. It penalizes the sum of absolute value (ℓ_1 -norm) of weights by regression analysis and shrinks some coefficients with the others reset to 0. The informative features are obtained by ridge regression and subset selection. Elastic Net [37] is proposed by Zou *et al.* as a regularized regression method. It contains the main penalty function built by ℓ_1 and ℓ_2 penalties of the Lasso and ridge methods. Similar to Lasso, the elastic net simultaneously carries out variable selection automatically and shrinkage continuously and it can choose the highly relevant features.

Multi-Modal Multi-Task (M3T) [38] method comes without regularization terms that trains a selection model to obtain a joint subset which consists of common relevant features using multi-task feature selection from each modality for multiple response variables. It noteworthy that, M3T is a special case of the proposed method by setting $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 0$.

Zhu *et al.* [39] (<https://sites.google.com/site/seanzhuxf/>) method simultaneously conducts classification and clinical scores prediction based on an improved loss function considering element-wise similarity and variable-variable relations.

Multi-modal sparse learning (MMSL) [22] method obtained from Lei *et al.* simultaneously performs classification and clinical scores prediction based on an improved loss function that considers the relations among rows and the information among columns in response variables.

C. Performance Evaluation

To estimate the performance, we utilize several quantitative measurements including accuracy (ACC), sensitivity (SEN), precision (PREC), specificity (SPEC), F-Score, and area under the receiver operating characteristic (ROC) curve (AUC). To validate the effectiveness of regression between the predicted and target clinical scores, we further calculate the correlation coefficient (CC) and root mean squared error (RMSE).

D. Classification Performance

Tables II, III, and IV summarize the classification results of NC vs. PD, NC vs. SWEDD, and PD vs. SWEDD from single modality features to multi-modal features. We also compare the ROC curves of the competing methods on these binary classification problems in Fig. 2. From these results, we have the following findings.

First, it is more effective to use the joint multi-task feature selection method to perform simultaneous classification and clinical scores prediction (M3T, MMSL, and our method) than the separate classification and regression model (Lasso and Elastic method). Taking NC vs. SWEDD in GM data as an example, M3T achieves an ACC of 87.67%, an F-Score of 91%, and an AUC of 89.98% and MMSL achieves an ACC of 90.78%, an F-Score of 94.24%, and AUC of 96.31% while the proposed method achieves an ACC of 93.89%, an F-Score of 95.45%, and an AUC of 96.37%. However, Lasso only achieves an ACC of 85.44%, an F-Score of 89.74%, and an AUC of 88.53% whereas Elastic method merely has an ACC of 83.44%, an F-Score of 88.43%, and an AUC of 88.95%.

Second, MD data is effective in improving the classification performances in NC vs. PD and PD vs. SWEDD. The possible reason is that, MD data is more comprehensive and microcosmic to reflect the brain information. For example, our method has the best performance with MD data compared to other modal data in NC vs. PD and gains an ACC of 91.57%, an F-Score of 85.63% and an AUC of 94.78%. Meanwhile, in PD vs. SWEDD, our method obtains an ACC of 92.06%, an F-Score of 95.39% and an AUC of 94.63%. As for NC vs. SWEDD, our method with CSF data compared with MD data has a slightly high ACC value (e.g., 93.89% vs. 91.89%).

Third, the regularization terms (e.g., multiple relationships among responses, features, and subjects) are effective in obtaining disease-related features. For example, in GCM, compared to M3T without the regularization terms, the accuracy in our method with the regularization terms increase by 4.91% for discriminating NC and PD, 4.33% for discriminating NC and SWEDD, and 2.88% for discriminating PD and SWEDD.

TABLE II
CLASSIFICATION PERFORMANCE OF VARIOUS METHODS IN NC VS. PD. BOLDFACE DENOTES BEST PERFORMANCE (MEAN $\pm$ STANDARD DEVIATION).

Feature	Method	NC vs. PD						
		ACC(%)	SEN(%)	SPEC(%)	PREC(%)	F1-score (%)	AUC(%)	p-value
GM	Lasso	80.83 $\pm$ 7	64.52 $\pm$ 17	88.00 $\pm$ 9	72.93 $\pm$ 19	66.73 $\pm$ 3	86.62 $\pm$ 8	<0.001
	Elastic Net	83.33 $\pm$ 5	64.52 $\pm$ 17	91.67 $\pm$ 8	81.48 $\pm$ 17	69.32 $\pm$ 11	87.34 $\pm$ 7	0.002
	M3T	86.19 $\pm$ 6	70.48 $\pm$ 11	92.95 $\pm$ 6	84.21 $\pm$ 15	75.60 $\pm$ 9	88.31 $\pm$ 9	0.02
	Zhu <i>et al.</i>	85.74 $\pm$ 4	78.81 $\pm$ 8	94.48 $\pm$ 6	85.81 $\pm$ 16	76.13 $\pm$ 10	92.18 $\pm$ 5	0.04
	MMSL	88.52 $\pm$ 11	80.48$\pm$17	94.48$\pm$6	86.62$\pm$14	80.95 $\pm$ 17	92.18 $\pm$ 5	>0.05
	Proposed	89.07$\pm$7	78.57 $\pm$ 14	94.29 $\pm$ 6	85.67 $\pm$ 14	81.14$\pm$13	92.22$\pm$7	--
CSF	Lasso	82.31 $\pm$ 5	64.29 $\pm$ 20	90.10 $\pm$ 10	79.73 $\pm$ 17	67.43 $\pm$ 12	83.51 $\pm$ 8	<0.001
	Elastic Net	82.33 $\pm$ 5	65.48 $\pm$ 20	89.33 $\pm$ 7	75.73 $\pm$ 11	67.93 $\pm$ 12	83.95 $\pm$ 7	<0.001
	M3T	86.21 $\pm$ 5	68.81 $\pm$ 17	93.67 $\pm$ 6	85.19 $\pm$ 15	74.25 $\pm$ 11	90.92 $\pm$ 7	>0.05
	Zhu <i>et al.</i>	84.69 $\pm$ 6	75.00 $\pm$ 18	89.43 $\pm$ 7	76.19 $\pm$ 16	73.99 $\pm$ 11	89.27 $\pm$ 6	0.03
	MMSL	85.67 $\pm$ 7	75.48$\pm$12	92.19 $\pm$ 7	82.25 $\pm$ 18	74.36 $\pm$ 11	90.29 $\pm$ 5	0.03
	Proposed	86.71$\pm$4	75.24 $\pm$ 14	94.43$\pm$5	87.19$\pm$12	75.58$\pm$11	91.29$\pm$6	--
MD	Lasso	84.26 $\pm$ 7	75.71 $\pm$ 16	87.95 $\pm$ 10	76.65 $\pm$ 15	74.37 $\pm$ 10	89.87 $\pm$ 7	<0.001
	Elastic Net	83.79 $\pm$ 7	74.29 $\pm$ 22	88.00 $\pm$ 10	76.44 $\pm$ 15	72.46 $\pm$ 13	88.18 $\pm$ 8	<0.001
	M3T	85.21 $\pm$ 8	78.81 $\pm$ 15	88.00 $\pm$ 9	76.32 $\pm$ 17	76.41 $\pm$ 13	90.34 $\pm$ 5	<0.001
	Zhu <i>et al.</i>	87.64 $\pm$ 8	83.33 $\pm$ 21	91.57 $\pm$ 7	81.83 $\pm$ 18	79.52 $\pm$ 14	91.93 $\pm$ 8	0.01
	MMSL	89.17 $\pm$ 5	85.00 $\pm$ 17	94.38 $\pm$ 5	86.25 $\pm$ 12	82.18 $\pm$ 12	93.90 $\pm$ 6	>0.05
	Proposed	90.10$\pm$7	91.90$\pm$12	96.48$\pm$4	91.00$\pm$10	83.89$\pm$11	94.85$\pm$6	--
GCM	Lasso	85.29 $\pm$ 7	72.38 $\pm$ 16	90.81 $\pm$ 9	80.94 $\pm$ 15	74.78 $\pm$ 11	91.01 $\pm$ 7	<0.001
	Elastic Net	84.79 $\pm$ 6	70.95 $\pm$ 20	90.90 $\pm$ 6	78.88 $\pm$ 13	72.95 $\pm$ 12	90.50 $\pm$ 7	<0.001
	M3T	86.21 $\pm$ 6	78.81 $\pm$ 14	94.29 $\pm$ 6	85.67 $\pm$ 14	76.92 $\pm$ 10	91.43 $\pm$ 8	<0.001
	Zhu <i>et al.</i>	88.64 $\pm$ 6	78.57 $\pm$ 16	95.00 $\pm$ 6	87.50 $\pm$ 14	80.28 $\pm$ 10	92.60 $\pm$ 6	0.009
	MMSL	89.17 $\pm$ 5	81.90 $\pm$ 18	95.71 $\pm$ 5	89.25 $\pm$ 13	81.66 $\pm$ 10	93.92 $\pm$ 7	0.02
	Proposed	91.12$\pm$7	91.67$\pm$12	96.52$\pm$5	91.57$\pm$11	85.55$\pm$11	94.88$\pm$5	--

TABLE III
CLASSIFICATION PERFORMANCE OF VARIOUS METHODS IN NC vs. SWEDD. BOLDFACE DENOTES BEST PERFORMANCE (MEAN $\pm$ STANDARD DEVIATION).

Feature	Method	NC vs. SWEDD						
		ACC(%)	SEN(%)	SPEC(%)	PREC(%)	F1-score (%)	AUC(%)	p-value
GM	Lasso	85.44 $\pm$ 10	96.67 $\pm$ 7	65.00 $\pm$ 25	84.69 $\pm$ 10	89.74 $\pm$ 6	88.53 $\pm$ 16	0.01
	Elastic Net	83.44 $\pm$ 10	96.67 $\pm$ 7	60.83 $\pm$ 25	82.55 $\pm$ 11	88.43 $\pm$ 6	88.95 $\pm$ 13	0.002
	M3T	87.67 $\pm$ 6	100.00 $\pm$ 0	70.83 $\pm$ 13	85.95 $\pm$ 6	91.00 $\pm$ 7	89.98 $\pm$ 7	0.002
	Zhu <i>et al.</i>	88.67 $\pm$ 9	100.00 $\pm$ 0	74.17 $\pm$ 19	87.62 $\pm$ 10	91.84 $\pm$ 7	91.79 $\pm$ 8	0.01
	MMSL	90.78 $\pm$ 7	100.00 $\pm$ 0	75.83 $\pm$ 24	89.64 $\pm$ 10	94.24 $\pm$ 4	96.31 $\pm$ 6	0.05
	Proposed	93.89$\pm$5	100.00$\pm$0	85.00$\pm$16	93.21$\pm$7	95.45$\pm$4	96.37$\pm$5	--
CSF	Lasso	85.44 $\pm$ 8	96.67 $\pm$ 11	62.50 $\pm$ 30	85.04 $\pm$ 10	89.63 $\pm$ 6	94.56 $\pm$ 8	<0.001
	Elastic Net	85.44 $\pm$ 10	100.00 $\pm$ 0	57.50 $\pm$ 29	82.77 $\pm$ 10	90.29 $\pm$ 6	89.15 $\pm$ 15	<0.001
	M3T	88.67 $\pm$ 10	100.00 $\pm$ 0	70.00 $\pm$ 28	87.42 $\pm$ 12	92.24 $\pm$ 7	94.37 $\pm$ 9	0.01
	Zhu <i>et al.</i>	90.78 $\pm$ 10	100.00 $\pm$ 0	77.50 $\pm$ 17	89.56 $\pm$ 10	93.67 $\pm$ 7	95.62 $\pm$ 6	0.05
	MMSL	91.89 $\pm$ 10	100.00 $\pm$ 0	84.17 $\pm$ 23	93.49 $\pm$ 9	94.42 $\pm$ 5	95.06 $\pm$ 7	>0.05
	Proposed	93.89$\pm$9	100.00$\pm$0	90.83$\pm$22	96.35$\pm$8	95.67$\pm$5	98.10$\pm$5	--
MD	Lasso	83.44 $\pm$ 7	95.24 $\pm$ 8	63.33 $\pm$ 25	83.39 $\pm$ 11	88.16 $\pm$ 5	88.10 $\pm$ 4	<0.001
	Elastic Net	81.44 $\pm$ 8	96.67 $\pm$ 7	54.17 $\pm$ 25	80.44 $\pm$ 10	87.19 $\pm$ 5	90.46 $\pm$ 8	<0.001
	M3T	86.67 $\pm$ 7	96.67 $\pm$ 7	72.50 $\pm$ 21	86.90 $\pm$ 10	90.13 $\pm$ 5	93.81 $\pm$ 6	0.01
	Zhu <i>et al.</i>	87.44 $\pm$ 5	100.00 $\pm$ 0	72.50 $\pm$ 19	87.86 $\pm$ 7	92.48 $\pm$ 3	95.16 $\pm$ 5	0.02
	MMSL	88.78 $\pm$ 7	100.00 $\pm$ 0	77.50 $\pm$ 17	88.81 $\pm$ 8	91.84 $\pm$ 6	95.26 $\pm$ 6	0.05
	Proposed	91.89$\pm$8	100.00$\pm$0	81.67$\pm$16	91.55$\pm$7	94.17$\pm$6	97.64$\pm$3	--
GCM	Lasso	87.67 $\pm$ 6	98.33 $\pm$ 5	67.50 $\pm$ 17	85.22 $\pm$ 7	91.14 $\pm$ 5	93.67 $\pm$ 8	0.001
	Elastic Net	87.56 $\pm$ 9	98.33 $\pm$ 5	66.67 $\pm$ 28	85.87 $\pm$ 9	91.40 $\pm$ 6	88.93 $\pm$ 13	<0.001
	M3T	90.56 $\pm$ 8	100.00 $\pm$ 0	75.00 $\pm$ 26	89.56 $\pm$ 10	93.34 $\pm$ 5	94.98 $\pm$ 8	0.02
	Zhu <i>et al.</i>	91.78 $\pm$ 6	100.00 $\pm$ 0	77.50 $\pm$ 22	90.40 $\pm$ 9	94.13 $\pm$ 4	94.58 $\pm$ 10	0.05
	MMSL	91.89 $\pm$ 9	100.00 $\pm$ 0	82.50 $\pm$ 22	91.59 $\pm$ 9	94.24 $\pm$ 5	96.23 $\pm$ 6	0.05
	Proposed	94.89$\pm$7	100.00$\pm$0	88.33$\pm$22	94.92$\pm$8	96.30$\pm$5	97.80$\pm$3	--

TABLE IV
CLASSIFICATION PERFORMANCE OF VARIOUS METHODS IN PD vs. SWEDD. BOLDFACE DENOTES BEST PERFORMANCE (MEAN $\pm$ STANDARD DEVIATION).

Feature	Method	PD vs. SWEDD						
		ACC(%)	SEN(%)	SPEC(%)	PREC(%)	F1-score (%)	AUC(%)	p-value
GM	Lasso	86.90 $\pm$ 6	99.29 $\pm$ 2	33.33 $\pm$ 33	86.92 $\pm$ 6	92.55 $\pm$ 3	87.56 $\pm$ 10	>0.05
	Elastic Net	86.96 $\pm$ 5	97.86 $\pm$ 3	40.00 $\pm$ 39	88.19 $\pm$ 7	92.49 $\pm$ 3	87.92 $\pm$ 9	0.04
	M3T	88.10 $\pm$ 3	99.29 $\pm$ 2	40.83 $\pm$ 35	88.15 $\pm$ 7	93.09 $\pm$ 2	90.20 $\pm$ 5	>0.05
	Zhu <i>et al.</i>	88.07 $\pm$ 6	99.29 $\pm$ 2	43.33 $\pm$ 36	88.73 $\pm$ 7	93.15 $\pm$ 3	87.68 $\pm$ 10	>0.05
	MMSL	88.66$\pm$6	100.00 $\pm$ 0	48.33 $\pm$ 33	89.62 $\pm$ 6	93.51$\pm$3	92.45 $\pm$ 6	>0.05
	Proposed	88.63$\pm$5	100.00$\pm$0	48.33$\pm$31	89.62$\pm$5	93.36$\pm$3	92.45$\pm$6	--
CSF	Lasso	87.52 $\pm$ 5	100.00 $\pm$ 0	35.00 $\pm$ 27	86.82 $\pm$ 5	92.87 $\pm$ 3	84.61 $\pm$ 14	>0.05
	Elastic Net	88.14 $\pm$ 5	99.29 $\pm$ 2	42.50 $\pm$ 24	87.87 $\pm$ 5	93.13 $\pm$ 3	83.73 $\pm$ 12	>0.05
	M3T	88.69 $\pm$ 7	99.29 $\pm$ 2	48.33 $\pm$ 37	89.16 $\pm$ 7	93.45 $\pm$ 4	90.96 $\pm$ 8	>0.05
	Zhu <i>et al.</i>	88.73 $\pm$ 5	100.00 $\pm$ 0	42.50 $\pm$ 24	87.94 $\pm$ 5	93.50 $\pm$ 3	89.67 $\pm$ 9	>0.05
	MMSL	89.28 $\pm$ 7	100.00 $\pm$ 0	48.33 $\pm$ 37	89.28 $\pm$ 7	93.82 $\pm$ 4	91.08$\pm$7	>0.05
	Proposed	89.28$\pm$7	100.00$\pm$0	48.33$\pm$34	89.28$\pm$7	93.82$\pm$4	89.67 $\pm$ 9	--
MD	Lasso	88.69 $\pm$ 5	97.86 $\pm$ 3	50.83 $\pm$ 34	89.72 $\pm$ 7	93.38 $\pm$ 2	86.05 $\pm$ 11	0.005
	Elastic Net	88.66 $\pm$ 5	99.29 $\pm$ 2	44.17 $\pm$ 25	88.36 $\pm$ 5	93.41 $\pm$ 3	83.83 $\pm$ 11	0.004
	M3T	89.84 $\pm$ 4	99.29 $\pm$ 2	50.83 $\pm$ 20	89.45 $\pm$ 5	94.03 $\pm$ 3	87.15 $\pm$ 7	0.04
	Zhu <i>et al.</i>	89.84 $\pm$ 4	100.00 $\pm$ 0	51.67 $\pm$ 29	90.57 $\pm$ 5	94.31 $\pm$ 3	90.32 $\pm$ 7	0.04
	MMSL	91.57 $\pm$ 5	100.00 $\pm$ 0	63.33 $\pm$ 25	91.81 $\pm$ 6	94.98 $\pm$ 3	91.01 $\pm$ 7	>0.05
	Proposed	92.06$\pm$6	100.00$\pm$0	65.83$\pm$36	93.22$\pm$7	95.39$\pm$4	94.63$\pm$5	--
GCM	Lasso	89.84 $\pm$ 4	99.29 $\pm$ 2	51.67 $\pm$ 18	89.45 $\pm$ 5	94.03 $\pm$ 3	86.39 $\pm$ 12	0.03
	Elastic Net	89.28 $\pm$ 4	99.29 $\pm$ 2	48.33 $\pm$ 18	88.90 $\pm$ 4	93.73 $\pm$ 2	86.22 $\pm$ 11	0.01
	M3T	89.84 $\pm$ 4	100.00 $\pm$ 0	48.33 $\pm$ 29	89.08 $\pm$ 6	94.12 $\pm$ 3	87.91 $\pm$ 8	0.03
	Zhu <i>et al.</i>	91.50 $\pm$ 4	100.00 $\pm$ 0	59.17 $\pm$ 17	91.14 $\pm$ 4	94.98 $\pm$ 2	89.66 $\pm$ 10	>0.05
	MMSL	92.12$\pm$5	100.00 $\pm$ 0	60.00 $\pm$ 21	91.27 $\pm$ 5	95.37$\pm$3	92.72 $\pm$ 8	>0.05
	Proposed	92.12 $\pm$ 6	100.00$\pm$0	65.83$\pm$30	92.62$\pm$7	95.34 $\pm$ 3	93.82$\pm$6	--

TABLE V
Comparison of our proposed method using different settings of regularization parameters in NC vs. PD. BOLDFACE IS BEST PERFORMANCE ACHIEVED (MEAN $\pm$ STANDARD DEVIATION)

Regularization parameters	NC vs. PD		NC vs. SWEDD		PD vs. SWEDD	
	ACC(%)	AUC(%)	ACC(%)	AUC(%)	ACC(%)	AUC(%)
$\lambda_1=\lambda_2=0, \lambda_3, \lambda_4, \lambda_5>0$	87.24 $\pm$ 8	90.95 $\pm$ 7	91.89 $\pm$ 6	95.42 $\pm$ 9	91.50 $\pm$ 4	88.54 $\pm$ 11
$\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5>0$	91.12$\pm$7	94.88$\pm$5	94.89$\pm$7	97.80$\pm$3	92.12$\pm$6	93.82$\pm$6

Fig. 2. Comparison of ROC curves for the competing methods on GM, CSF, MD, and GCM (Top: NC vs. PD, middle: NC vs. SWEDD, and bottom: PD vs. SWEDD)

Fourth, by majority voting, GCM effectively improves the classification performance compared with single modal data. For example, compared with single modal data, the proposed method has the best performance with multi-modal data and obtains an ACC of 91.12%, a SEN of 91.67%, a PREC of 91.57%, an F-Score of 85.55%, and an AUC of 84.4% in NC vs. PD. In the NC vs. SWEDD experiment, the proposed method achieves an ACC of 94.89%, a SEN of 100.00%, a PREC of 94.92%, an F-Score of 96.30%, and an AUC of 97.80%. In the PD vs. SWEDD experiment, we have an ACC of 92.12%, a SEN of 100.00%, a PREC of 92.62%, an F-Score of 95.34%, and an AUC of 93.82%. Fig. 3 also shows classification performance for the competing methods with GCM data. We can observe that our method achieves the best performances compared with other methods. Meanwhile, Table V verifies that the first two regularization terms cannot be ignored.

Although the ACC values of our method and MMSL are very close and similar in the PD and SWEDD (no significant difference), both methods have the best classification performance compared with other methods. For example, in CSF data, the proposed method and MMSL obtain the same ACC, SEN, SPEC, PREC, and F1-score of 89.28%, 100%, 48.33%, 89.28%, and 93.82%, respectively. In GCM data, the proposed method has a slightly higher standard deviation (e.g., 6% vs. 5%) than MMSL under the same ACC of 92.12%.

Finally, to evaluate the statistical significance of two AUCs, we conduct Delong's test to compare the performances from different methods (95% significance level). In GM data, our method can significantly outperform both Lasso and Elastic net methods. Also, in CSF data, our method can significantly

outperform Elastic net method for NC vs. SWEDD. In MD data, there is significant difference between our method and Elastic net method. In GCM data, those joint multi-task methods (i.e., M3T, Zhu *et al.*'s method, MMSL, and our method) have significant difference compared with Lasso for NC vs. SWEDD. Meanwhile, the proposed method can significantly outperform Lasso, Elastic net, M3T, and Zhu *et al.*'s methods for PD vs. SWEDD. Although there almost are no significant differences between the two AUCs from different methods, our method has higher AUCs than the other methods.

E. Regression Performance

The values of CC and RMSE are used to evaluate the performance of the regression model. Tables VI, VII, and VIII show the regression performances of the competing methods in terms of NC vs. PD, NC vs. SWEDD, and PD vs. SWEDD, respectively. We can see that our method has the best regression performance.

In NC vs. PD, our method achieves the best performance in the prediction of depression, sleep, olfaction, and MoCA scores. The best performance with MD using CC and RMSE is 0.583 and 1.284 in depression score, 0.564 and 3.213 in sleep score, 0.573 and 7.806 in olfaction score, and 0.643 and 1.613 in MoCA score. In NC vs. SWEDD, our proposed method has the best performance to predict depression and MoCA scores. The best performance with CSF is 0.820 (CC) and 0.886 (RMSE) in depression score, and 0.840 (CC) and 1.278 (RMSE) in MoCA score. MMSL achieves the best performance in sleep and olfaction scores. The best performance with CSF is 0.816 (CC) and 3.019 (RMSE) in sleep score, and 0.873 (CC) and 4.276 (RMSE) in olfaction score.

Fig. 3. Comparison of classification performances for the competing methods on GCM data.

As for PD vs. SWEDD, our proposed method obtains the best performance in prediction of depression, sleep, olfaction, and MoCA scores. The best performance with MD using CC and RMSE are 0.600 and 1.304, 0.561 and 3.279, and 0.589 and 7.688 in depression, sleep, and olfaction, respectively. Meanwhile, we can also see that although prediction performance of our method and MMSL is similar in olfaction score, our method has slightly higher prediction performance from single modal data to multi-modal data, such as 0.589 (CC) vs. 0.582 (CC) in MD data and 0.585 (CC) vs. 0.583 (CC) in GCM data. In MoCA score, the best performance with GCM is 0.662 (CC) and 1.843 (RMSE). Fig. 4 also shows regression performance for the competing methods that have the best results selected from GM, CSF, MD, and GCM. We can see that our method has the best performance. In the meantime, to

check statistical significance, we implement the paired t -tests (95% significance level) on regression performances of our method and the competing methods, Tables VI, VII and VIII show the p -values computed from the values of CC. These resulting p -values indicate that our method is statistically superior to the competing methods on the tasks of predicting clinical scores.

F. Results Summary

From our experimental results, we observe that GCM data can better improve classification and regression performances in NC vs. PD, PD vs. SWEDD, and NC vs. SWEDD. The proposed method with multi-modal features is far preferable to all other methods in the classification tasks.

Meanwhile, we illustrate the importance of different brain regions using histogram and exploit BrainNet Viewer [40] to visualize the regional importance on the brain in Fig. 6. The top ten brain regions selected from NC vs. PD classification with GCM data using our proposed method are precentral gyrus left, precuneus left, caudate nucleus left, heschl gyrus left, temporal pole: middle temporal gyrus left, superior temporal gyrus left, heschl gyrus right, temporal pole: superior temporal gyrus right, middle temporal gyrus left, and inferior temporal gyrus left. For NC vs. SWEDD classification, the most informative features are precentral gyrus left, middle frontal gyrus left, insula left, anterior cingulate and paracingulate gyri left, rolandic operculum right, inferior frontal gyrus right, superior frontal gyrus left, superior occipital gyrus, parahippocampal gyrus right, and median cingulate and paracingulate gyri left. For PD vs. SWEDD classification, the most informative features are precentral gyrus right, middle occipital gyrus right, median cingulate and paracingulate gyri right, superior frontal gyrus right, posterior cingulate gyrus, gyrus rectus left, gyrus rectus right, rolandic operculum right, middle frontal gyrus left, and insula left. For example, in [41], Yu *et al.* found that the blood oxygenation level dependent activation in the motor cortex located in the posterior precentral gyrus immediately anterior to the central sulcus was positively correlated with the severity of upper limb rigidity in PD. In the presence of external regulatory rhythmical stimuli (music beats), PD patients' gait has gained improvement [42] and heschl gyrus and superior temporal gyrus are related with auditory sense. Meanwhile, in [43-45], caudate nucleus, middle frontal gyrus, and middle occipital gyrus are related with PD. In these binary classifications, the top most brain region is precentral gyrus in all scenarios, which indicates that this brain area may be closely related to PD. Finally, to further find other brain regions related to top 10 brain regions, we first compute the Pearson correlation coefficient matrix using the weight coefficient matrix of 10-fold cross-validation in the loss function. Then, we find the five most relevant brain regions to each top 10 regions by the matrix. Finally, we use the *circularGraph* function of MATLAB to draw the correlation maps. Fig. 5 shows other brain regions that are related to top10 brain regions, which may be helpful for further investigation on PD in the future.

Fig. 4. Regression performance of all competing methods that have the best results selected from GM, CSF, MD, and GCM in terms of CC for depression, sleep, olfaction, and MoCA scores prediction.

Fig. 5. Importance of different brain regions and associated visual representation on GCM data. (top) NC and PD, (middle) NC and SWEDD, and (bottom) PD and SWEDD (Note that; these vivid visual representations are generated by projecting the importance to cerebral cortex. The value of the importance has been normalized to the interval [1, 255]. The larger the value, the more important the corresponding brain region).

To further find other brain regions related to top 10 brain regions, we first compute the Pearson correlation coefficient matrix using the weight coefficient matrix of 10-fold cross-validation on the loss function. Then, we find the five most relevant brain regions to each top 10 regions by the matrix. Finally, we use the *circularGraph* function of MATLAB to draw the correlation maps as shown in Fig. 6. In this figure, we observe that there are seven brain regions strongly associated with the top 10 discriminative brain regions in NC vs. PD, without considering the top 10 discriminative brain areas themselves. These regions are posterior cingulate gyrus left, lingual gyrus left, lingual gyrus right, inferior occipital gyrus left, inferior occipital gyrus right, superior temporal gyrus right, and temporal pole of middle temporal gyrus right. In NC vs.

SWEDD, correlation maps of feature weights between sensitive ROIs and other regions are complex and scattered, which may be one of the reasons why SWEDD is difficult to find the causative factor. For PD vs. SWEDD, we can see that there are seven brain regions in the top10 brain regions that are very closely correlated. They are Rolandic operculum right, medial orbital superior frontal gyrus right, gyrus rectus left, gyrus rectus right, insula left, middle cingulum right, and posterior cingulate gyrus left. We believe these findings can be helpful for further investigations on PD diagnosis in the future.

V. DISCUSSIONS

Although the proposed method achieves the appealing performance, there are still some limitations. First, AAL atlas used to extract features in this paper only has 116 ROIs segmented from brain. In fact, multiple-scale templates used to extract features in [46] has excellent classification performance. The reason may be that multiple-scale templates may extract more comprehensive features from coarse-grained to fine-grained scales. Second, our work only uses MRI and DTI to conduct ensemble classification and regression. Since multi-modal data can reflect brain's information from different points, we can use more modalities such as positron emission tomography (PET) and fMRI for more robust ensemble classification and regression. Third, in our work, we use SVM as classifier. However, it is possible that other classifiers may achieve better performance such as extreme learning machine [47]. Finally, the features extracted in this work are hand-crafted biomarkers, which are quite limited. In fact, convolution neural network can automatically learn the discriminative features, which seems feasible to boost the diagnosis performance [48]. In future, we can further improve our classification and regression performance by addressing these limitations.

VI. CONCLUSIONS

In this paper, we propose a joint multi-task feature selection framework to simultaneously conduct classification and regression for PD diagnosis from multimodal data. Instead of using single modality, multiple modalities are explored, which enhances the feature representation abilities. Multiple relationships among features, subjects, and clinical scores are utilized to discover the internal and inherent information for

performance boosting. By constructing a united and multi-task joint learning framework, our proposed method can identify the most disease-related biomarkers, which is beneficial for PD

monitoring. Our extensive experiments based on PPMI dataset show that the performance of the proposed method outperforms its counterparts.

Fig. 6. Correlation maps of feature weights between sensitive ROIs marked with red fonts and other ROIs denoted with black fonts (Left: NC vs. PD, middle: NC vs. SWEDD, and right: PD vs. SWEDD).

TABLE V
REGRESSION PERFORMANCE OF VARIOUS METHODS IN NC VS. PD. BOLDFACE IS BEST PERFORMANCE (MEAN $\pm$ STANDARD DEVIATION).

		NC vs. PD								
Feature	Method	Depression scores		Sleep scores		Olfaction scores		MoCA scores		p
		CC	RMSE	CC	RMSE	CC	RMSE	CC	RMSE	
GM	Lasso	0.461±0.09	1.355±0.30	0.490±0.07	3.374±0.60	0.471±0.15	8.594±1.10	0.509±0.14	1.675±0.57	0.007
	Elastic	0.437±0.08	1.367±0.29	0.486±0.06	3.375±0.60	0.476±0.14	8.565±1.09	0.522±0.11	1.683±0.56	<0.001
	M3T	0.479±0.14	1.332±0.32	0.498±0.07	3.373±0.59	0.463±0.16	8.581±1.09	0.536±0.10	1.669±0.56	0.02
	Zhu <i>et al.</i>	0.518±0.09	1.328±0.31	0.498±0.07	3.373±0.59	0.533±0.27	8.524±1.05	0.536±0.10	1.653±0.53	0.05
	MMSL	0.518±0.09	1.293±0.25	0.518±0.08	3.349±0.60	0.533±0.27	8.244±0.81	0.554±0.12	1.644±0.52	>0.05
	Proposed	0.524±0.08	1.291±0.30	0.509±0.06	3.339±0.58	0.505±0.16	8.395±1.10	0.552±0.10	1.656±0.55	--
CSF	Lasso	0.468±0.06	1.366±0.28	0.509±0.09	3.364±0.66	0.455±0.10	8.459±1.24	0.517±0.05	1.724±0.49	0.001
	Elastic	0.467±0.07	1.375±0.28	0.491±0.11	3.362±0.66	0.503±0.09	8.405±1.22	0.491±0.08	1.703±0.48	<0.001
	M3T	0.462±0.09	1.350±0.29	0.499±0.10	3.337±0.62	0.519±0.10	8.276±1.12	0.508±0.13	1.689±0.46	0.001
	Zhu <i>et al.</i>	0.527±0.07	1.345±0.28	0.532±0.09	3.339±0.69	0.519±0.10	8.276±1.12	0.520±0.11	1.714±0.48	0.008
	MMSL	0.528±0.08	1.337±0.27	0.538±0.08	3.306±0.60	0.541±0.07	8.249±1.14	0.526±0.08	1.668±0.41	>0.05
	Proposed	0.532±0.06	1.327±0.26	0.552±0.08	3.308±0.54	0.515±0.10	8.269±1.13	0.541±0.09	1.681±0.47	--
MD	Lasso	0.522±0.08	1.352±0.30	0.471±0.09	3.387±0.59	0.485±0.09	8.260±1.09	0.530±0.14	1.725±0.44	<0.001
	Elastic	0.486±0.14	1.346±0.31	0.459±0.10	3.398±0.56	0.468±0.08	8.262±1.08	0.570±0.09	1.740±0.46	<0.001
	M3T	0.512±0.12	1.316±0.29	0.496±0.09	3.304±0.59	0.521±0.06	8.193±1.01	0.570±0.09	1.707±0.38	0.003
	Zhu <i>et al.</i>	0.533±0.09	1.296±0.30	0.506±0.09	3.330±0.64	0.518±0.09	8.069±1.06	0.573±0.10	1.695±0.42	0.001
	MMSL	0.562±0.08	1.296±0.30	0.532±0.10	3.200±0.52	0.566±0.14	7.777±1.02	0.582±0.09	1.629±0.39	0.003
	Proposed	0.577±0.08	1.284±0.29	0.564±0.10	3.213±0.74	0.573±0.08	7.806±0.97	0.643±0.11	1.613±0.43	--
GCM	Lasso	0.463±0.13	1.355±0.30	0.455±0.10	3.362±0.53	0.484±0.10	8.232±1.05	0.534±0.14	1.716±0.44	<0.001
	Elastic	0.452±0.14	1.355±0.30	0.442±0.09	3.404±0.55	0.487±0.11	8.235±1.00	0.564±0.11	1.724±0.47	<0.001
	M3T	0.504±0.15	1.329±0.29	0.469±0.09	3.319±0.63	0.505±0.12	8.097±0.95	0.560±0.11	1.701±0.47	0.002
	Zhu <i>et al.</i>	0.510±0.12	1.317±0.29	0.484±0.07	3.342±0.67	0.521±0.13	7.996±0.81	0.545±0.14	1.685±0.40	0.002
	MMSL	0.532±0.10	1.311±0.30	0.500±0.11	3.228±0.61	0.543±0.08	7.915±0.96	0.565±0.10	1.633±0.39	0.003
	Proposed	0.542±0.12	1.277±0.27	0.532±0.10	3.175±0.57	0.570±0.10	7.917±1.04	0.621±0.09	1.612±0.45	--

TABLE VI
REGRESSION PERFORMANCE OF VARIOUS METHODS IN NC VS. SWEDD. BOLDFACE IS BEST PERFORMANCE (MEAN $\pm$ STANDARD DEVIATION).

Feature	Method	NC vs. SWEDD								p
		Depression scores		Sleep scores		Olfaction scores		MoCA scores		
		CC	RMSE	CC	RMSE	CC	RMSE	CC	RMSE	
GM	Lasso	0.691 ±0.15	1.140 ±0.47	0.689 ±0.13	3.678 ±0.89	0.641 ±0.15	5.369 ±2.19	0.659 ±0.17	1.660 ±0.76	<0.001
	Elastic	0.671 ±0.15	1.090 ±0.43	0.692 ±0.12	3.709 ±0.90	0.633 ±0.17	5.354 ±2.22	0.676 ±0.16	0.655 ±0.71	<0.001
	M3T	0.704 ±0.15	1.085 ±0.51	0.725 ±0.16	3.641 ±0.99	0.724 ±0.12	5.203 ±1.87	0.676 ±0.17	1.584 ±0.57	<0.001
	Zhu <i>et al.</i>	0.732 ±0.13	0.960 ±0.46	0.776 ±0.12	3.392 ±0.92	0.713 ±0.12	4.839 ±1.54	0.716 ±0.16	1.513 ±0.45	0.003

	MMSL	0.753±0.09	0.938±0.41	0.799±0.10	3.470±0.63	0.788±0.10	4.786±1.62	0.767±0.17	1.413±0.37	0.004
	Proposed	0.810±0.08	0.947±0.46	0.790±0.12	3.179±0.62	0.835±0.19	4.695±1.98	0.811±0.08	1.339±0.31	--
	Lasso	0.694±0.13	1.026±0.39	0.660±0.12	3.531±0.90	0.682±0.15	4.742±1.86	0.704±0.11	1.599±0.81	<0.001
	Elastic	0.709±0.15	1.083±0.41	0.705±0.14	3.379±0.84	0.750±0.13	4.506±1.90	0.708±0.12	1.588±0.74	<0.001
	M3T	0.731±0.11	0.940±0.35	0.713±0.08	3.306±0.62	0.749±0.12	4.399±1.59	0.768±0.19	1.490±0.73	<0.001
	Zhu <i>et al.</i>	0.749±0.11	0.922±0.29	0.752±0.10	3.266±0.55	0.771±0.11	4.468±1.81	0.759±0.10	1.430±0.45	<0.001
	MMSL	0.793±0.11	0.897±0.35	0.816±0.10	3.019±0.77	0.873±0.47	4.276±1.56	0.752±0.09	1.449±0.58	0.02
	Proposed	0.820±0.09	0.886±0.35	0.786±0.09	3.052±0.80	0.839±0.25	3.882±1.65	0.840±0.29	1.278±0.45	--
	Lasso	0.660±0.10	1.084±0.46	0.715±0.14	3.540±1.04	0.732±0.14	4.727±1.67	0.709±0.15	1.423±0.61	0.001
	Elastic	0.637±0.09	1.107±0.48	0.733±0.14	3.455±1.11	0.734±0.14	4.778±1.91	0.713±0.17	1.381±0.57	<0.001
	M3T	0.681±0.13	1.067±0.43	0.727±0.12	3.450±1.20	0.724±0.12	4.557±1.79	0.746±0.16	1.414±0.59	<0.001
	Zhu <i>et al.</i>	0.753±0.09	1.019±0.40	0.740±0.15	3.323±0.72	0.761±0.10	4.325±1.65	0.735±0.15	1.375±0.61	0.003
	MMSL	0.737±0.11	1.014±0.42	0.771±0.12	3.335±1.22	0.784±0.08	4.426±1.65	0.753±0.14	1.363±0.46	0.04
	Proposed	0.755±0.17	0.925±0.23	0.789±0.12	3.209±0.47	0.796±0.08	4.271±1.68	0.797±0.20	1.326±0.59	--
	Lasso	0.688±0.12	1.027±0.42	0.642±0.14	3.544±0.92	0.688±0.16	4.735±1.84	0.693±0.14	1.559±0.81	<0.001
	Elastic	0.697±0.17	1.087±0.44	0.709±0.14	3.456±0.80	0.766±0.12	4.431±1.85	0.709±0.13	1.531±0.72	<0.001
	M3T	0.732±0.12	0.931±0.34	0.722±0.07	3.333±0.66	0.752±0.14	4.417±1.59	0.728±0.16	1.475±0.68	<0.001
	Zhu <i>et al.</i>	0.758±0.09	0.900±0.29	0.747±0.10	3.278±0.32	0.778±0.11	4.111±1.55	0.724±0.10	1.471±0.69	<0.01
	MMSL	0.774±0.12	0.891±0.32	0.771±0.15	3.194±0.84	0.806±0.14	4.076±1.74	0.784±0.11	1.438±0.53	>0.05
	Proposed	0.797±0.08	0.854±0.21	0.796±0.13	3.091±0.24	0.820±0.11	4.194±1.79	0.788±0.14	1.302±0.60	--

TABLE VII

REGRESSION PERFORMANCE OF VARIOUS METHODS IN PD VS. SWEDD. BOLDFACE IS BEST PERFORMANCE (MEAN±STANDARD DEVIATION).

		PD vs. SWEDD								
Feature	Method	Depression scores		Sleep scores		Olfaction scores		MoCA scores		p
		CC	RMSE	CC	RMSE	CC	RMSE	CC	RMSE	
GM	Lasso	0.535±0.10	1.411±0.32	0.447±0.17	3.525±0.73	0.471±0.09	8.545±0.73	0.500±0.12	2.023±0.56	0.001
	Elastic	0.498±0.10	1.398±0.32	0.484±0.10	3.530±0.72	0.484±0.10	8.572±0.90	0.511±0.11	2.024±0.55	<0.001
	M3T	0.521±0.07	1.372±0.29	0.511±0.09	3.528±0.71	0.507±0.07	8.609±0.85	0.553±0.09	2.027±0.55	0.03
	Zhu <i>et al.</i>	0.517±0.08	1.352±0.32	0.526±0.10	3.467±0.80	0.470±0.09	8.521±0.89	0.554±0.14	1.957±0.58	>0.05
	MMSL	0.541±0.07	1.352±0.32	0.529±0.13	3.467±0.80	0.512±0.02	8.359±0.88	0.560±0.15	1.950±0.53	>0.05
	Proposed	0.528±0.14	1.352±0.32	0.536±0.11	3.467±0.80	0.512±0.02	8.387±0.79	0.556±0.08	1.950±0.53	--
CSF	Lasso	0.459±0.07	1.422±0.31	0.490±0.15	3.549±0.75	0.513±0.16	8.163±0.52	0.570±0.10	2.021±0.66	0.001
	Elastic	0.460±0.08	1.401±0.28	0.525±0.13	3.513±0.74	0.544±0.13	8.312±0.54	0.582±0.08	2.009±0.65	<0.001
	M3T	0.536±0.10	1.373±0.32	0.489±0.13	3.507±0.64	0.517±0.14	8.272±0.52	0.557±0.09	2.007±0.65	<0.001
	Zhu <i>et al.</i>	0.548±0.11	1.315±0.32	0.539±0.10	3.490±0.69	0.540±0.15	8.001±0.73	0.604±0.07	1.974±0.63	>0.05
	MMSL	0.573±0.09	1.280±0.25	0.540±0.10	3.454±0.73	0.541±0.15	8.001±0.73	0.605±0.10	1.946±0.58	0.03
	Proposed	0.549±0.11	1.315±0.32	0.540±0.10	3.496±0.79	0.541±0.15	8.001±0.73	0.604±0.07	1.975±0.63	--
MD	Lasso	0.492±0.11	1.401±0.33	0.471±0.10	3.417±0.83	0.544±0.10	8.082±0.78	0.569±0.08	1.981±0.61	<0.001
	Elastic	0.483±0.07	1.394±0.33	0.484±0.07	3.371±0.76	0.552±0.14	8.109±0.84	0.578±0.09	1.975±0.62	<0.001
	M3T	0.529±0.09	1.358±0.31	0.498±0.08	3.392±0.80	0.563±0.11	8.015±0.81	0.576±0.08	1.980±0.61	<0.001
	Zhu <i>et al.</i>	0.544±0.08	1.335±0.31	0.523±0.10	3.326±0.80	0.567±0.10	7.956±0.82	0.610±0.09	1.919±0.64	0.008
	MMSL	0.576±0.12	1.321±0.29	0.530±0.07	3.317±0.65	0.582±0.11	7.836±1.24	0.615±0.17	1.890±0.61	0.05
	Proposed	0.600±0.10	1.304±0.30	0.561±0.09	3.279±0.81	0.589±0.11	7.688±1.11	0.620±0.10	1.867±0.55	--
GCM	Lasso	0.476±0.13	1.400±0.33	0.454±0.10	3.437±0.80	0.561±0.09	8.168±0.75	0.596±0.08	1.955±0.61	<0.001
	Elastic	0.497±0.08	1.395±0.33	0.440±0.09	3.404±0.75	0.566±0.13	8.145±0.80	0.598±0.07	1.953±0.63	<0.001
	M3T	0.499±0.10	1.352±0.25	0.464±0.11	3.380±0.72	0.574±0.09	8.054±0.84	0.593±0.07	1.955±0.60	<0.001
	Zhu <i>et al.</i>	0.519±0.11	1.359±0.34	0.499±0.07	3.349±0.79	0.580±0.11	7.954±0.91	0.590±0.08	1.922±0.60	<0.001
	MMSL	0.547±0.14	1.321±0.31	0.516±0.14	3.302±0.69	0.583±0.09	7.924±1.33	0.609±0.10	1.871±0.59	>0.05
	Proposed	0.574±0.12	1.276±0.29	0.537±0.07	3.266±0.77	0.585±0.13	7.662±1.28	0.622±0.05	1.843±0.60	--

REFERENCES

- [1] L. V. Kalia and A. E. Lang, "Parkinson's disease," *The Lancet*, vol. 386, pp. 896-912, Aug 2015.
- [2] R. B. Postuma, D. Berg, M. Stern, W. Poewe, C. W. Olanow, W. Oertel, *et al.*, "MDS clinical diagnostic criteria for Parkinson's disease," *Mov Disord*, vol. 30, pp. 1591-601, Oct 2015.
- [3] Y. Kim, S.-M. Cheon, C. Youm, M. Son, and J. W. Kim, "Depression and posture in patients with Parkinson's disease," *Gait & Posture*, vol. 61, pp. 81-85, Mar 2018.
- [4] D. W. Dickson, "Neuropathology of Parkinson disease," *Parkinsonism & Related Disorders*, vol. 46, pp. S30-S33, Jan 2018.
- [5] J. Lotharius and P. Brundin, "Pathogenesis of parkinson's disease: dopamine, vesicles and [alpha]-synuclein," *Nat Rev Neurosci*, vol. 3, pp. 932-942, Jan 2002.
- [6] P. Schwingenschuh, D. Ruge, M. J. Edwards, C. Terranova, P. Katschnig, F. Carrillo, *et al.*, "Distinguishing SWEDDs patients with asymmetric resting tremor from Parkinson's disease: a clinical and electrophysiological study," *Mov Disord*, vol. 25, pp. 560-9, Apr 2010.
- [7] N. Nicastro, V. Garibotto, S. Badoud, and P. R. Burkhard, "Scan without evidence of dopaminergic deficit: A 10-year retrospective study," *Parkinsonism & Related Disorders*, vol. 31, pp. 53-58, Oct 2016.
- [8] I. Obeso, E. Casabona, R. Rodriguez-Rojas, M. L. Bringas, R. Macias, N. Pavon, *et al.*, "Unilateral subthalamotomy in Parkinson's disease: Cognitive, psychiatric and neuroimaging changes," *Cortex*, vol. 94, pp. 39-48, Sep 2017.
- [9] R. Martínez-Fernández, R. Rodríguez-Rojas, M. del Álamo, F. Hernández-Fernández, J. A. Pineda-Pardo, M. Dileone, *et al.*, "Focused ultrasound subthalamotomy in patients with asymmetric Parkinson's disease: a pilot study," *The Lancet Neurology*, vol. 17, pp. 54-63, Jan 2018.
- [10] A. L. Benabid, S. Chabardes, J. Mitrofanis, and P. Pollak, "Deep brain stimulation of the subthalamic nucleus for the treatment of Parkinson's disease," *Lancet Neurol*, vol. 8, pp. 67-81, Jan 2009.
- [11] G. Singh and L. Samavedham, "Unsupervised learning based feature extraction for differential diagnosis of neurodegenerative diseases: A case study on early-stage diagnosis of Parkinson disease," *J Neurosci Methods*, vol. 256, pp. 30-40, Dec 2015.

- [12] E. Adeli, F. Shi, L. An, C.-Y. Wee, G. Wu, T. Wang, *et al.*, "Joint feature-sample selection and robust diagnosis of Parkinson's disease from MRI data," *NeuroImage*, vol. 141, pp. 206-219, Nov 2016.
- [13] R. Prashanth, S. D. Roy, P. K. Mandal, and S. Ghosh, "Automatic classification and prediction models for early Parkinson's disease diagnosis from SPECT imaging," *Expert Systems with Applications*, vol. 41, pp. 3333-3342, Jun 2014.
- [14] A. Abós, H. C. Baggio, B. Segura, A. I. Garc á-D áz, Y. Compta, M. J. Martí *et al.*, "Discriminating cognitive status in Parkinson's disease through functional connectomics and machine learning," *Scientific Reports*, vol. 7, p. 45347, Mar 2017.
- [15] L. Liu, Q. Wang, E. Adeli, L. Zhang, H. Zhang, and D. Shen, "Exploring diagnosis and imaging biomarkers of Parkinson's disease via iterative canonical correlation analysis based feature selection," *Computerized Medical Imaging and Graphics*, vol. 67, pp. 21-29, Jul 2018.
- [16] B. Lei, F. Jiang, S. Chen, D. Ni, and T. Wang, "Longitudinal Analysis for Disease Progression via Simultaneous Multi-Relational Temporal-Fused Learning," *Frontiers in Aging Neuroscience*, vol. 9, Mar 2017.
- [17] E. Adeli, G. Wu, B. Saghafi, L. An, F. Shi, and D. Shen, "Kernel-based Joint Feature Selection and Max-Margin Classification for Early Diagnosis of Parkinson's Disease," *Scientific Reports*, vol. 7, p. 41069, Jan 2017.
- [18] E. Adeli, K. H. Thung, L. An, G. Wu, F. Shi, T. Wang, *et al.*, "Semi-Supervised Discriminative Classification Robust to Sample-Outliers and Feature-Noises," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1-1, 2018.
- [19] F. J. A. Meijer, B. R. Bloem, P. Mähknecht, K. Seppi, and B. Goraj, "Update on diffusion MRI in Parkinson's disease and atypical parkinsonism," *Journal of the Neurological Sciences*, vol. 332, pp. 21-29, Sep 2013.
- [20] H. I. Suk, S. W. Lee, and D. Shen, "Subclass-based multi-task learning for Alzheimer's disease diagnosis," *Front Aging Neurosci*, vol. 6, p. 168, 2014.
- [21] B. Lei, P. Yang, T. Wang, S. Chen, and D. Ni, "Relational-Regularized Discriminative Sparse Learning for Alzheimer's Disease Diagnosis," *IEEE Trans Cybern*, vol. 47, pp. 1102-1113, Apr 2017.
- [22] H. Lei, Z. Huang, J. Zhang, Z. Yang, E.-L. Tan, F. Zhou, *et al.*, "Joint detection and clinical score prediction in Parkinson's disease via multi-modal sparse learning," *Expert Systems with Applications*, vol. 80, pp. 284-296, Sep 2017.
- [23] X. Zhu, H.-I. Suk, L. Wang, S.-W. Lee, and D. Shen, "A novel relational regularization feature selection method for joint regression and classification in AD diagnosis," *Medical Image Analysis*, vol. 38, pp. 205-214, May 2017.
- [24] J. A. Yesavage, "Geriatric Depression Scale," *Psychopharmacol Bull*, vol. 24, pp. 709-11, 1988.
- [25] M. W. Johns, "A new method for measuring daytime sleepiness: the Epworth sleepiness scale," *Sleep*, vol. 14, pp. 540-5, Dec 1991.
- [26] R. L. Doty, P. Shaman, C. P. Kimmelman, and M. S. Dann, "University of Pennsylvania Smell Identification Test: a rapid quantitative olfactory function test for the clinic," *Laryngoscope*, vol. 94, pp. 176-8, Feb 1984.
- [27] N. Kandiah, A. Zhang, A. R. Cenina, W. L. Au, N. Nadkarni, and L. C. Tan, "Montreal Cognitive Assessment for the screening and prediction of cognitive decline in early Parkinson's disease," *Parkinsonism Relat Disord*, vol. 20, pp. 1145-8, Nov 2014.
- [28] K. Marek, D. Jennings, S. Lasch, A. Siderowf, C. Tanner, T. Simuni, *et al.*, "The Parkinson Progression Marker Initiative (PPMI)," *Progress in Neurobiology*, vol. 95, pp. 629-635, Dec 2011.
- [29] S. A. Sadananthan, W. Zheng, M. W. Chee, and V. Zagorodnov, "Skull stripping using graph cuts," *Neuroimage*, vol. 49, pp. 225-39, Jan 2010.
- [30] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, *et al.*, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *Neuroimage*, vol. 15, pp. 273-89, Jan 2002.
- [31] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, "FSL," *Neuroimage*, vol. 62, pp. 782-790, Aug 2012.
- [32] G. F. Wooten, L. J. Currie, V. E. Bovbjerg, J. K. Lee, and J. Patrie, "Are men at greater risk for Parkinson's disease than women?," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 75, p. 637, 2004.
- [33] G. E. Gillies, I. S. Pienaar, S. Vohra, and Z. Qamhawi, "Sex differences in Parkinson's disease," *Frontiers in Neuroendocrinology*, vol. 35, pp. 370-384, 2014.
- [34] J. C. Jurado-Coronel, R. Cabezas, M. F. Ávila Rodríguez, V. Echeverría, L. M. Garc á-Segura, and G. E. Barreto, "Sex differences in Parkinson's disease: Features on clinical symptoms, treatment outcome, sexual hormones and genetics," *Frontiers in Neuroendocrinology*, Sep 2017.
- [35] A. Reeve, E. Simcox, and D. Turnbull, "Ageing and Parkinson's disease: Why is advancing age the biggest risk factor?," *Ageing Research Reviews*, vol. 14, pp. 19-30, Mar 2014.
- [36] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society Series B-Methodological*, vol. 58, pp. 267-288, 1996.
- [37] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society Series B-Statistical Methodology*, vol. 67, pp. 301-320, 2005.
- [38] D. Zhang and D. Shen, "Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in alzheimer's disease," *NeuroImage*, vol. 59, pp. 895-907, 2012.
- [39] X. Zhu, H. I. Suk, and D. Shen, "A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis," *Neuroimage*, vol. 100, pp. 91-105, Oct 2014.
- [40] M. Xia, J. Wang, and Y. He, "BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics," *PLOS ONE*, vol. 8, p. e68910, 2013.
- [41] H. Yu, D. Sternad, D. M. Corcos, and D. E. Vaillancourt, "Role of Hyperactive Cerebellum and Motor Cortex in Parkinson's Disease," *NeuroImage*, vol. 35, pp. 222-233, Jan 2007.
- [42] C. Nombela, L. E. Hughes, A. M. Owen, and J. A. Grahn, "Into the groove: can rhythm influence Parkinson's disease?," *Neurosci Biobehav Rev*, vol. 37, pp. 2564-70, Dec 2013.
- [43] B. Lach, D. Grimes, B. Benoit, and A. Minkiewicz-Janda, "Caudate nucleus pathology in Parkinson's disease: ultrastructural and biochemical findings in biopsy material," *Acta Neuropathol*, vol. 83, pp. 352-60, 1992.
- [44] H. Watanabe, J. Senda, S. Kato, M. Ito, N. Atsuta, K. Hara, *et al.*, "Cortical and subcortical brain atrophy in Parkinson's disease with visual hallucination," *Mov Disord*, vol. 28, pp. 1732-6, Oct 2013.
- [45] L. Sang, J. Zhang, L. Wang, J. Zhang, Y. Zhang, P. Li, *et al.*, "Alteration of Brain Functional Networks in Early-Stage Parkinson's Disease: A Resting-State fMRI Study," *PLoS One*, vol. 10, p. e0141815, 2015.
- [46] M. Liu, D. Zhang, and D. Shen, "Relationship Induced Multi-Template Learning for Diagnosis of Alzheimer's Disease and Mild Cognitive Impairment," *IEEE transactions on medical imaging*, vol. 35, pp. 1463-1474, Jun 2016.
- [47] H.-L. Chen, G. Wang, C. Ma, Z.-N. Cai, W.-B. Liu, and S.-J. Wang, "An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson's disease," *Neurocomputing*, vol. 184, pp. 131-144, Apr 2016.
- [48] M. Liu, J. Zhang, D. Nie, P. T. Yap, and D. Shen, "Anatomical Landmark based Deep Feature Representation for MR Images in Brain Disease Diagnosis," *IEEE Journal of Biomedical and Health Informatics*, vol. PP, pp. 1-1, 2018.