A Multi-Task Deep Feature Selection Method for Brain Imaging Genetics

Chenglin Yu, Shu Zhang, Muheng Shang, Lei Guo, Junwei Han, and Lei Du

Abstract—Using brain imaging quantitative traits (QTs) for identifying genetic risk factors is an important research topic in brain imaging genetics. Many efforts have been made for this task via building linear models between imaging QTs and genetic factors such as single nucleotide polymorphisms (SNPs). To the best of our knowledge, linear models could not fully uncover the complicated relationship due to the loci’s elusive and diverse influences on imaging QTs. In this paper, we propose a novel multi-task deep feature selection (MTDFS) method for brain imaging genetics. MTDFS first builds a multi-task deep neural network to model the complicated associations between imaging QTs and SNPs. And then designs a multi-task one-to-one layer and imposes a combined penalty to identify SNPs that make significant contributions. MTDFS can not only extract the nonlinear relationship but also arms the deep neural network with feature selection. We compared MTDFS to multi-task linear regression (MTLR) and single-task DFS (DFS) methods on the real neuroimaging genetic data. The experimental results showed that MTDFS performed better than MTLR and DFS on the QT-SNP relationship identification and feature selection. Thus, MTDFS is powerful for identifying risk loci and could be a great supplement to brain imaging genetics.

Index Terms—Brain imaging genetics, deep feature selection, multi-task deep feature selection, multi-task learning.

I. INTRODUCTION

RECENTLY, brain imaging genetics attracts more and more attention owing to its superior power in identifying genetic risk factors compared to the case-control studies [1], [2], [3]. In this area, the imaging quantitative traits (QTs) and single nucleotide polymorphisms (SNPs) are usually analyzed jointly, with aim to reveal novel risk loci for brain disorders such as Alzheimer’s disease (AD) [2], [4] and schizophrenia (SZ) [5], [6], [7].

Up to now, there have been many efforts made for imaging genetics. For example, based on the univariate linear method, Shen et al. [8] used the structural brain imaging measures as imaging QTs and confirmed several risk loci showing relevance to AD. Wang et al. [9] used SNPs to predict imaging QTs based on the multi-task linear regression (MTLR). This method found a link between risk loci and several pre-selected imaging QTs that were attacked by AD. Sparse canonical correlation analysis (SCCA) was also used to explore the relationship between imaging QTs and SNPs, which usually combined with the feature selection techniques to identify risk loci [10], [11], [12], [13], [14], [15], [16], [17], [18]. A common issue of these methods is that they are linear models and thus may be insufficient to reveal the complicated yet challenging mechanism of the heritability of human brain, as a result of that genetic factors could hardly follow a linear relationship to affect the brain structure and function, as well as brain disorders [19], [20]. More importantly, the generation of brain imaging phenotypes requires a series of processes such as DNA transcription, splicing, and translation, which are difficult to consistently follow linear patterns. So it is difficult for linear models to explain the complex QT-SNP relationship.

Recently, deep neural network (DNN) has shown great success in many applications such as image classification and objective detection [21], [22]. The DNN can extract nonlinear relationship between imaging QTs and SNPs, but it usually suffers from the interpretation issue. In other word, a conventional DNN model cannot tell us that which SNPs, usually conceal in a large candidate SNP set, contribute significantly to the imaging QTs. Therefore, it is essential to explore novel DNN models with good interpretable ability, which has the potential to identify meaningful loci that linear models may miss.

In this article, we designed a multi-task deep neural network method with feature selection (MTDFS) to study the nonlinear correlation between SNPs and brain imaging QTs, and simultaneously to identify SNPs of relevance. First, MTDFS builds a multi-task DNN based prediction model where SNPs were independent variables and imaging QTs were dependent variables. This strategy could model the nonlinear relationship between SNPs and QTs, which may yield interesting relationship that missed by linear models. Second, to figure out SNPs of relevance, MTDFS introduced a one-to-one layer in front of the multi-task DNN, and imposed sparsity-inducing penalties on the multi-task one-to-one layer. This setup implemented the
feature selection for SNPs, and thus made the MTDFS easy to interpret.

We used real neuroimaging genetic data obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database, and compared MTDFS to one multi-task linear method (MTLR for short) [9] and one single-task DFS (DFS for short) [23]. The results showed that MTDFS held lower root mean square (RMSE) and higher correlation coefficient (CC) than both MTLR and DFS. Interestingly, MTDFS also obtained better regression weight profiles because that most of its identified SNPs were related to abnormal imaging QTs, and further to AD. In contrast, MTLR reported too many relevant SNPs which was hard to interpret, and DFS’s weights were also denser than MTDFS which means it had lower performance than MTDFS. In summary, MTDFS had superior performance to both MTLR and DFS, showing that it could be very promising for brain imaging genetics. In addition, we also conducted ablation experiments on the combined penalty. Compared with the common combination of $\ell_{1,1}$-norm, $\ell_{2,1}$-norm and $G_{2,1}$-norm [9], [24], our proposed regularization can give MTDFS better performance and generalization ability. Moreover, this combined penalty can also identify different types of genetic factors, including single variant gene locus, gene loci shared by multiple tasks, and multiple variant gene loci of same linkage disequilibrium (LD) block. In summary, the experimental results indicate that the proposed model has significant research value in brain imaging genetics.

II. METHODS

A. Background

Using brain imaging QTs as intermediated phenotypes has shown great success in identifying genetic risk loci. In generally, brain areas are pre-selected, being called regions of interest (ROI), and their neuroimaging measurements such as the grey matter density are then extracted as quantitative traits (QTs). After that, a single-task regression or multi-task regression model is built to predict these imaging QTs using SNP data as independent variables [9]. Since not all SNPs are responsible for brain disorders, the regularization techniques are employed to select those SNPs of relevance [10], [11], [12], [15], [16], [17]. However, these methods can only find out the linear relationship owing to the linear modeling technique. As analyzed earlier, the human genome could nonlinearly affect the brain structure and function via nonlinearly influencing those abnormal imaging QTs. DNN could be a desirable alternative, but a critical issue is that it cannot select features for the input space. Li et al. introduced a DNN with feature selection ability, mainly by adding a sparse one-to-one layer before the input layer of the DNN [23]. But this model only applies to single task, which indicates that it ignores the relationship among multiple interrelated tasks, i.e., predicting multiple correlated imaging QTs based on SNPs in this paper.

B. The MTDFS Model

To extract the nonlinear relationship between multiple imaging QTs and SNPs, and with aim to identify relevant SNPs, we propose the multi-task deep feature selection (MTDFS) method.

MTDFS builds a multi-task DNN model to predict multiple interrelated imaging QTs based on a common set of SNPs, and simultaneously adds a multi-task one-to-one layer with aim to identify SNPs showing relevance to these QTs. The framework of MTDFS is presented in Fig. 1. There are two components, i.e., the nonlinear relationship extraction component (red box) and the feature selection component (blue box). The nonlinear relationship extraction component extracts the nonlinear relationship hierarchically, and then employs the multi-task regression to predict multiple interrelated imaging QTs. The feature selection component uses a sparse multi-task one-to-one layer in front of the multi-task DNN, and imposes penalties on this layer to ensure that only relevant SNPs can be fed into the nonlinear relationship extraction component. Therefore, MTDFS can not only extract nonlinear relationships, but also select features of interest.

By convention, we denote scalars as italic letters, column vectors as boldface lowercase letters, and matrices as boldface capitals. $\mathbf{X} \in \mathbb{R}^{p \times n}$ denotes the genetic data with $n$ subjects and $p$ SNPs. $\mathbf{Y} \in \mathbb{R}^{n \times T}$ represents the brain imaging data, where $T$ is the number of imaging QTs (tasks). According to Fig. 1, we use $\mathbf{U} \in \mathbb{R}^{p \times T}$ denote the weights of the multi-task one-to-one layer. The $i$th row of $\mathbf{U}$ is denoted as $\mathbf{U}_i$. $\mathbf{W}$ and $\mathbf{b}$ are the weights and bias of the deep neural network, respectively.

Then MTDFS is formally written as,

$$\min_{\mathbf{U}, \mathbf{W}, \mathbf{b}} \mathcal{L}(\mathbf{U}, \mathbf{W}, \mathbf{b}) + \mathcal{R}(\mathbf{U}),$$

where $\mathcal{L}(\mathbf{U}, \mathbf{W}, \mathbf{b})$ is the loss function for the nonlinear relationship extraction component and $\mathcal{R}(\mathbf{U})$ is the regularization term corresponding to the feature selection component. Next, we will present both components in details.

1) Nonlinear Relationship Extraction: The nonlinear relation extraction component is essentially a multi-layer perceptron network. In the MTDFS model, we further denote parameters of the $k$th layer as $\mathbf{W}^{(k)}$ and $\mathbf{b}^{(k)}$ where $k$ represents the layer of interest. When $k$ is the last layer, then it represents the output layer and those remaining ones correspond to hidden layers.
Now $\mathcal{L}(U, W, b)$ can be written as,

$$
\mathcal{L}(U, W, b) = l \left( U, W^{(1)}, b^{(1)}, \ldots, W^{(K)}, b^{(K)} \right),
$$

(2)

where $l(U, W^{(1)}, b^{(1)}, \ldots, W^{(K)}, b^{(K)})$ represents the DNN objective, and $K$ is the total number of the layers. Since we have multiple interrelated imaging QTs corresponding to multiple tasks, these $U$’s, $W$’s and $b$’s will be jointly optimized following the multi-task learning.

As shown in Fig. 1, the $i$th task in the weighted input layer can be expressed by $H_i^{(0)} = X_i \odot U_i$, where $\odot$ is Hadamard product. Thus we have

$$
H_i^{(0)} = \begin{bmatrix} X_1 \odot U_1 & \cdots & X_i \odot U_i & \cdots & X_T \odot U_T \end{bmatrix}.
$$

(3)

Now these $H_i^{(0)}$’s are fed into the DNN via sharing the same set of parameters $W$.

Then for each hidden layer, we have

$$
H_i^{(j)} = \sigma \left( W^{(j)} H_i^{(j-1)} + b^{(j)} \right)
$$

(4)

associating with the $i$th task, where $\sigma$ is the activation function such as the sigmoid in this work. $H_i^{(j)}$ is the output of the $(j-1)$th layer and the input of the $(j+1)$th layer.

From Fig. 1 and (4), we know that parameters for all hidden layers are shared. Finally we arrive at

$$
\hat{y}_i = W_i^{(k)} H_i^{(k-1)} + b^{(k-1)}
$$

(5)

with $\hat{y}_i$ denoting the $i$th predicted imaging QT. We use the regression function in this paper since imaging QTs are continuous. Certainly, other prediction methods such as the negative log-likelihood (NLL) or softmax function can also be used if applicable.

Finally, we can write $\mathcal{L}(U, W, b)$ as follows,

$$
\mathcal{L}(U, W, b) = \| Y - \hat{Y} \|_F^2 = \sum_{i=1}^T \| y_i - \hat{y}_i \|_2^2,
$$

(6)

where $y_i$ is the true value of the $i$th imaging QT, $Y$ and $\hat{Y}$ are the matrix form of all $y_i$ and $\hat{y}_i$. During the optimization, the objective $\mathcal{L}(U, W, b)$ will be jointly optimized with the feature selection component which will shown below.

2) Feature Selection: To select a feature subset out of the whole input space, we employ a multi-task one-to-one layer with regularization and add it to the front of the conventional DNN. It is worth noting that our model has better generalization ability than DFS since MTDFS will reduce to DFS when there is only one task [23]. As shown in Fig. 1, the regularization techniques are used for this additional layer. Suppose there are $T$ QTs (tasks), we will have $T$ one-to-one layers, with each corresponding to one imaging QT (task). To enable a reasonable model, we here are interested in three distinct types of sparsity, i.e., the element-sparsity, individual-sparsity, and group-sparsity shown in Fig. 2 [17]. Specifically, denoting the weight for the multi-task one-to-one layer as $U$, the regularization terms are defined as follows,

$$
R(U) = \lambda \| U \|_{G_{2,1}} + \beta \| U \|_{2,1} + \gamma \| U \|_{1,1},
$$

(7)

where $\lambda$, $\beta$ and $\gamma$ are non-negative parameters which control the sparsity of multi-task one-to-one layer.

In (7), the $G_{2,1}$-norm indicates the group-sparsity, and can select a group of genetic variations in the same linkage disequilibrium (LD) block for multiple interrelated tasks. According to [9], the $G_{2,1}$-norm is written as follows,

$$
\| U \|_{G_{2,1}} = \sum_{m=1}^M \left( \sum_{t=1}^T \sum_{i \in g_m} (u_{it})^2 \right)^{1/2},
$$

(8)

where $M$ is the number of LD, and $g_m$ indicates the $m$th LD set. The group-sparsity, illustrated in Fig. 2, can select the SNPs shared among all tasks.

Meanwhile, due to the complicated genetic effects, not all of the SNP in an effecting LD are related to AD and a SNP in an unrelated LD may also be effective. We use both $\ell_{2,1}$-norm and $\ell_{1,1}$-norm in our model since the intangible genetic effects could happen to multiple QTs synergistically or a specific QT alone. The $\ell_{2,1}$-norm is formulated as follows,

$$
\| U \|_{2,1} = \sum_{i=1}^p \| u_i \|_2 = \sum_{i=1}^p \sqrt{ \sum_{t=1}^T (u_{it})^2 },
$$

(9)

where $p$ is the number of SNPs. This penalty help select a relevant feature for multiple interrelated tasks simultaneously. And we call it individual-sparsity since it can identify a single SNP that shared among multiple tasks [17]. Then element-sparsity is attained by the $\ell_{1,1}$-norm, i.e.,

$$
\| U \|_{1,1} = \sum_{i=1}^p \sum_{j=1}^q | u_{ij} | ,
$$

(10)

which investigates that whether a SNP is effective for a specific imaging QT.

III. EXPERIMENTS

A. Real Neuroimaging Genetic Data

The genotyping and brain imaging data used in this paper were downloaded from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). One primary goal of ADNI is to test whether information such as serial magnetic resonance imaging (MRI), positron emission tomography (PET),
other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer’s disease (AD). See www.adni-info.org for the latest information.

We downloaded the real data from the LONI website (adni.loni.usc.edu), including the 18-F fluorobetapir PET (AV45) scans, fluorodeoxyglucose PET (FDG) scans and genetic data. Table I contained the details of 755 non-Hispanic Caucasian participants including 281 AD, 292 MCI and 182 healthy control (HC).

The PET scans were preprocessed following the pipeline including averaging, alignment to a standard space, resampling to a standard image and voxel size, smoothness to a uniform resolution and normalization to a cerebellar gray matter reference region which finally yielded standardized uptake value ratio images [25]. Before experiments, we normalized these images to Montreal Neurological Institute (MNI) space to 8 mm³ images. From the MarsBar AAL atlas, we obtained amyloid measurements at the ROI level. Since this experiment is AD-oriented, we carefully selected ten (AD-related) ROI level imaging QTs in different PET scans [17], [26], [27], [28]. Table II shows the description of these QTs. In addition, we chose 2000 genetic variations of chromosome 19 which are near to the APOE gene. The LD information was pre-calculated. The goal was to evaluate whether the nonlinear relationship between these PET scans and SNPs was better than the linear one. In the meanwhile, we also aimed to identify the AD-risk SNPs during the nonlinear modeling.

**B. Implementation Details**

We compared MTDFS to two most related methods, i.e., MTLR [9] and DFS [23], to access the performance. MTLR employs a multi-task learning paradigm with the same hybrid sparsity-inducing penalty. It directly imposes three sparse penalties on coefficients of linear regression weights, while MTDFS imposes them to a multi-task one-to-one layer. Thus comparing MTDFS to MTLR could show whether the introduction of DNN is effective, thereby helping demonstrate the superior modeling capability of the nonlinear relationship extraction. DFS uses the single-task DNN to model the nonlinear relationship and an additional one-to-one layer to select features. Comparing to DFS could evaluate the performance of our proposed multi-task one-to-one layer and the feature selection ability of the hybrid sparsity-inducing penalty. Therefore, using both MTLR and DFS could fully evaluate our proposed MTDFS. Some other linear models such as sparse canonical correlation analysis [10], [16], [29] was excluded since they are a bilateral multivariate method and are not designed for prediction tasks. In other words, they were unsupervised methods and thus different from MTDFS. Moreover, Deep CCA [30] and Deep CCAE [31] were incapable of selecting relevant features related to the disease, and the results might be hard to explain. Therefore, they cannot be included in comparative studies either. The recent deep autoencoder-based G-MIND independently learned the representation of each modality, including imaging phenotypes and genotypes, and then combined them for disease classification [3]. G-MIND did not identify the associations between imaging data and genetic data, and thus was different to our method.

We here used ten imaging QTs and 2,000 SNPs, which resulted in ten interrelated tasks corresponding to ten sparse one-to-one layers. Therefore, the dimension of U was 2000 × 10, and the regularization terms was applied to U as defined in (7). There were 2,000 units in the weighted input layer. We used a two-layer DNN network for the nonlinear relationship extraction. The first hidden layer had 128 units and the second one had 64 units. In addition, We also use Sigmoid function as the activation of these two hidden layers. Since we had ten imaging QTs, the output layer had ten units accordingly, where each task corresponds to a unit. DFS employed the same DNN structure as MTDFS. The MTLR model utilized the multi-task regression with the same hybrid penalty as defined in (7). We used the five-fold cross-validation to seek suitable parameters with the candidate set [0.00001, 0.0001, 0.001, 0.01, 0.1]. Moreover, all methods used the same settings to ensure the fairness of experiments, including the data partition, number of iterations (1,000 in this paper) and operating environment (PyTorch 1.8.1).

### C. Improved Imaging Phenotype Prediction

To test the ability of nonlinear feature extraction, two evaluation criteria were used to access the performance of three methods. Since neuroimaging QTs were continuous, we first employed the popular root mean square error (RMSE), which is the smaller the better. Besides, we also utilized the correlation coefficient (CC) as another metric where higher values indicate better performance. We jointly evaluate the performance of these three methods based on both metrics. We presented both RMSEs and CCs in Tables III and IV. In both tables, we can clearly observe that our MTDFS obtained better scores than both MTLR and DFS in the testing set. Although MTLR achieved the best results in the training set, its test results were not good. Therefore, MTLR has overfitting phenomenon and cannot perform well in feature extraction. This indicated that MTDFS not only predicted the dependent imaging QTs with the smallest error, but also extracted higher relationship between imaging QTs and SNPs. In addition, the metrics of MTDFS are significantly better than DFS. Although DFS has better results in Task 7 and Task 8 in the fluorodeoxyglucose PET (FDG) scans, Fig. 3(b) shows that DFS do not accomplish our main task, Feature Selection. Giving up feature selection for higher CC and smaller RMSE is not what we want. This revealed that DNN coupled with multi-task feature selection could obtain improved imaging phenotype prediction and associations between SNPs and imaging QTs, demonstrating the success of MTDFS.

**TABLE I
PARTICIPANT CHARACTERISTICS**

|        | HC    | MCI   | AD    |
|--------|-------|-------|-------|
| Num    | 182   | 292   | 281   |
| Gender (M/F, %) | 48.90/51.10 | 48.63/51.37 | 53.38/46.62 |
| Handedness (R/L, %) | 89.56/10.44 | 88.70/11.30 | 90.39/9.61 |
| Age (mean±std) | 73.93±5.51 | 70.90±6.64 | 72.61±6.15 |
| Education (mean±std) | 16.43±2.68 | 16.18±2.68 | 15.95±2.82 |
### Table II

**Ten ROI Level Imaging QTS Selected in AV45 (left) and FDG (right)**

| AV45 (ROI and Regions) | FDG (ROI and Regions) |
|------------------------|-----------------------|
| Frontal_Inf_Orb_Left   | Amygdala_Left         |
| Frontal_Inf_Orb_Right  | Amygdala              |
| Frontal_Med_Orb_Left   | Angular_Left          |
| Frontal_Med_Orb_Right  | Angular_left          |
| Frontal_Mid_Left       | Angular_Right         |
| Frontal_Mid_Right      | Angular Right         |
| Frontal Sup_Medal_Left  | Anterior cingulate    |
| Frontal Sup_Medal_Right | and paracingulate     |
| Olfactory_Left         | Hippocampus           |
| Olfactory_Right        | Hippocampus           |
|                       | Occipital_Inf_Left    |
|                       | Occipital_left         |

**TABLE III**

**The Average RMSE Along With the Standard Deviation (In the Parentheses) of Ten Tasks for Two Scans, i.e., AV45 and FDG**

| Task 1     | Task 2     | Task 3     | Task 4     | Task 5     | Task 6     | Task 7     | Task 8     | Task 9     | Task 10    |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Training  |            |            |            |            |            |            |            |            |            |
| RMSE       |            |            |            |            |            |            |            |            |            |
| AV45       | MTLR       | 0.1380     | 0.1302     | 0.1615     | 0.1623     | 0.1435     | 0.1448     | 0.1387     | 0.1443     | 0.1316     |
|            | (0.005)    | (0.003)    | (0.008)    | (0.005)    | (0.003)    | (0.004)    | (0.003)    | (0.003)    | (0.003)    | (0.003)    |
|            | DFS        | 0.2137     | 0.1981     | 0.2601     | 0.2624     | 0.2267     | 0.2444     | 0.2182     | 0.2269     | 0.2026     |
|            | (0.007)    | (0.004)    | (0.008)    | (0.005)    | (0.004)    | (0.006)    | (0.007)    | (0.005)    | (0.006)    | (0.004)    |
|            | MTDPS      | 0.2123     | 0.1966     | 0.2586     | 0.2618     | 0.2253     | 0.2231     | 0.2163     | 0.2249     | 0.2060     |
|            | (0.006)    | (0.005)    | (0.008)    | (0.009)    | (0.005)    | (0.007)    | (0.006)    | (0.006)    | (0.006)    | (0.005)    |
| Testing   |            |            |            |            |            |            |            |            |            |
| RMSE       |            |            |            |            |            |            |            |            |            |
| FDG        | MTLR       | 0.2600     | 0.2410     | 0.3175     | 0.3203     | 0.2713     | 0.2705     | 0.2635     | 0.2741     | 0.2455     |
|            | (0.032)    | (0.022)    | (0.037)    | (0.040)    | (0.028)    | (0.029)    | (0.025)    | (0.027)    | (0.021)    | (0.018)    |
|            | DFS        | 0.2179     | 0.2024     | 0.2685     | 0.2712     | 0.2314     | 0.2295     | 0.2285     | 0.2315     | 0.2062     |
|            | (0.027)    | (0.025)    | (0.039)    | (0.027)    | (0.021)    | (0.022)    | (0.021)    | (0.022)    | (0.028)    | (0.025)    |
|            | MTDPS      | 0.2155     | 0.1998     | 0.2659     | 0.2683     | 0.2295     | 0.2274     | 0.2210     | 0.2293     | 0.2038     |
|            | (0.032)    | (0.022)    | (0.037)    | (0.040)    | (0.024)    | (0.022)    | (0.029)    | (0.030)    | (0.025)    | (0.022)    |

**TABLE IV**

**The Average CC Along With the Standard Deviation (In the Parentheses) of Ten Tasks for Two Scans, i.e., AV45 and FDG**

| Task 1     | Task 2     | Task 3     | Task 4     | Task 5     | Task 6     | Task 7     | Task 8     | Task 9     | Task 10    |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Training  |            |            |            |            |            |            |            |            |            |
| CC         |            |            |            |            |            |            |            |            |            |
| AV45       | MTLR       | 0.8135     | 0.8053     | 0.8392     | 0.8381     | 0.8268     | 0.8195     | 0.8269     | 0.8258     | 0.6108     |
|            | (0.019)    | (0.010)    | (0.012)    | (0.011)    | (0.006)    | (0.005)    | (0.010)    | (0.010)    | (0.011)    | (0.011)    |
|            | DFS        | 0.8932     | 0.8930     | 0.4648     | 0.4334     | 0.4198     | 0.4249     | 0.4352     | 0.4294     | 0.4001     |
|            | (0.015)    | (0.012)    | (0.017)    | (0.016)    | (0.008)    | (0.007)    | (0.014)    | (0.010)    | (0.017)    | (0.018)    |
|            | MTDPS      | 0.8984     | 0.8983     | 0.4529     | 0.4333     | 0.4247     | 0.4285     | 0.4426     | 0.4364     | 0.4081     |
|            | (0.016)    | (0.012)    | (0.012)    | (0.019)    | (0.017)    | (0.019)    | (0.017)    | (0.017)    | (0.013)    | (0.018)    |
| Testing   |            |            |            |            |            |            |            |            |            |
| CC         |            |            |            |            |            |            |            |            |            |
| FDG        | MTLR       | 0.2196     | 0.2307     | 0.2623     | 0.2669     | 0.2509     | 0.2421     | 0.2554     | 0.2519     | 0.2133     |
|            | (0.068)    | (0.076)    | (0.072)    | (0.069)    | (0.073)    | (0.084)    | (0.077)    | (0.080)    | (0.060)    | (0.061)    |
|            | DFS        | 0.3208     | 0.3234     | 0.3636     | 0.3451     | 0.3517     | 0.3530     | 0.3700     | 0.3652     | 0.3395     |
|            | (0.049)    | (0.067)    | (0.054)    | (0.062)    | (0.071)    | (0.078)    | (0.058)    | (0.063)    | (0.052)    | (0.056)    |
|            | MTDPS      | 0.3259     | 0.3289     | 0.3740     | 0.3546     | 0.3549     | 0.3561     | 0.3704     | 0.3694     | 0.3403     |
|            | (0.076)    | (0.083)    | (0.070)    | (0.063)    | (0.062)    | (0.071)    | (0.078)    | (0.067)    | (0.069)    | (0.064)    |

The best values in the testing set were shown in bold.
VI clearly shows that the RMSE results of combined respectively represent the RMSE and CC of MTDFS [38] [35] APOE [33] [34] (b) that our model selects many APOC1 APOE APOC1 and so forth, APOE ℓ [39] APOC1 and those of the o [42] NECTIN2 and those of the hybrid penalty, MTDFS also identified group structures, generalization ability in different data. In addition, owing to same SNPs in the two scans. This shows that MTDFS has good comparison of Fig. [42x188] e.g., SNPs of the the hybrid penalty, MTDFS also identified group structures, generalization ability in different data. In addition, owing to same SNPs in the two scans. This shows that MTDFS has good comparison of Fig. [42x212] owing to its nonlinear modeling. Therefore, it is essential to compare the feature selection results. We presented the heat map which exhibited the feature selection in Fig. 3. The most relevant features were highlighted in this figure. It was clear that MTLR reported too many relevant SNPs. This was hard to interpret since identifying too many markers provides little to no useful information. DFS alleviated the drawback of MTLR, but it still identified too many markers than our method. In particular, two tasks on the FDG data of DFS reported too many SNPs, while MTDFS was not. This demonstrated that learning multiple tasks independently was suboptimal for exploring the complicated relationship between SNPs and imaging QTs. MTDFS successfully identified a small subset out of the whole SNP candidate set. More importantly, the top identified SNPs of MTDFS, including rs6857 (NECTIN2) [32], rs769449 (APOE) [33], rs429358 (APOC1) [34], rs10414043 (APOC1) [35], rs7256200 (APOC1) [36], rs483082 (APOC1) [37], rs438811 (APOC1) [38], rs73052335 (APOC1) [39], rs12721051 (APOC1) [40], rs56131196 (APOC1) [41], rs4420638 (APOC1) [42], rs66626994 (APOC1P1) [33] and so forth, were all correlated to AD. Moreover, it can be seen from the comparison of Fig. 3(a) and (b) that our model selects many same SNPs in the two scans. This shows that MTDFS has good generalization ability in different data. In addition, owing to the hybrid penalty, MTDFS also identified group structures, e.g., SNPs of the APOE and those of the APOC1. These results demonstrated the success of our multi-task one-to-one layer, indicating that this strategy can endow a meaningful feature selection capability to the deep neural network.

D. Genetic Marker Identification

It is not surprising that DNN holds higher correlations than linear models owing to its nonlinear modeling. Therefore, it is essential to compare the feature selection results. We presented the heat map which exhibited the feature selection in Fig. 3. The most relevant features were highlighted in this figure. It was clear that MTLR reported too many relevant SNPs. This was hard to interpret since identifying too many markers provides little to no useful information. DFS alleviated the drawback of MTLR, but it still identified too many markers than our method. In particular, two tasks on the FDG data of DFS reported too many SNPs, while MTDFS was not. This demonstrated that learning multiple tasks independently was suboptimal for exploring the complicated relationship between SNPs and imaging QTs. MTDFS successfully identified a small subset out of the whole SNP candidate set. More importantly, the top identified SNPs of MTDFS, including rs6857 (NECTIN2) [32], rs769449 (APOE) [33], rs429358 (APOC1) [34], rs10414043 (APOC1) [35], rs7256200 (APOC1) [36], rs483082 (APOC1) [37], rs438811 (APOC1) [38], rs73052335 (APOC1) [39], rs12721051 (APOC1) [40], rs56131196 (APOC1) [41], rs4420638 (APOC1) [42], rs66626994 (APOC1P1) [33] and so forth, were all correlated to AD. Moreover, it can be seen from the comparison of Fig. 3(a) and (b) that our model selects many same SNPs in the two scans. This shows that MTDFS has good generalization ability in different data. In addition, owing to the hybrid penalty, MTDFS also identified group structures, e.g., SNPs of the APOE and those of the APOC1. These results demonstrated the success of our multi-task one-to-one layer, indicating that this strategy can endow a meaningful feature selection capability to the deep neural network.

E. Combined Penalty Analysis

To verify the performance of the proposed combined penalty, we used different regularization combinations on the MTDFS model, including ℓ_{1,1}-norm, ℓ_{2,1}-norm and G_{2,1}-norm. In addition, we compare it with common multi-task regularization [9], [24], and also use the popular root mean square error (RMSE) and correlation coefficient (CC) for evaluation criteria. Tables V and VI respectively represent the RMSE and CC of MTDFS in the two scans: the 18-F florbetapir PET (AV45) scans and fluorodeoxyglucose PET (FDG) scans. Similarly, we conducted a five-fold cross-validation based on different regularization, and used lower RMSE and higher CC as the experimental result. Table V clearly shows that the RMSE results of combined penalty in AV45 scan are consistent with those using only ℓ_{2,1}-norm, while in FDG scan they are the same as those using only ℓ_{1,1}-norm, and have the best performance in most of all tasks. This shows that the regularization that plays a key role in different data is not the same, and combined penalty can solve this problem. From Table VI, it can be seen that using only the ℓ_{1,1}-norm resulted in the model achieving higher correlation coefficients (CC) in certain tasks. Of course, it can also be observed that models using combined penalty also achieve better performance in most of all tasks. But overall, the proposed combined penalty has better generalization ability in all tasks. Therefore, combined penalty can not only improve the nonlinear relationship extraction ability of the model, but also identify risk gene loci at different levels, such as those unique to a single task, common to multiple tasks, and those under the same linkage disequilibrium (LD) block.

F. Imaging Genetic Correlation Interpretation

To better understand the identified associations, in Fig. 4, we presented the pairwise correlation between ten imaging QTs in these two scans and top twelve SNPs shared by all tasks, and results of the analysis of variance (ANOVA) showed that all values
We observed that rs429358 had the highest correlation values for all imaging QTs, showing its importance in predicting these prominent AD-altered brain areas. In addition, rs12721051, rs56131196, and rs4420638 held the same value, informing us that they had the same importance for AD prediction probably due to their high linkage disequilibrium relationship. The same results can be observed for rs10414043 and rs7256200, demonstrating the capability of our method in uncovering the group structure of human genome. Fig. 4 shows that SNPs selected by MTDFS on different data exhibited similar effects.

In addition, we can also see that rs429358 has a good correlation with the left orbital part of the middle frontal gyrus in the AV45 scans and left angular gyrus in the FDG scans, respectively. Based on this phenomenon, we conducted population stratification analysis using these two QTs to explore the genetic variation mechanisms of genes. Fig. 5(a) presents distributions of identified imaging QTs with rs429358 among different diagnostic groups, e.g., HCs, MCIs and ADs: (a) The left orbital part of middle frontal gyrus of the AV45 scan. (b) The left angular gyrus of the FDG scan.

The best values in the testing set were shown in bold.

| Task 1 | Task 2 | Task 3 | Task 4 | Task 5 | Task 6 | Task 7 | Task 8 | Task 9 | Task 10 |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 0.1583 | 0.2100 | 0.2432 | 0.2563 | 0.2449 | 0.2190 | 0.2155 | 0.2064 | 0.2079 | 0.2088 |

Fig. 5. The measurement distributions of identified imaging QTs with rs429358 among different diagnostic groups, e.g., HCs, MCIs and ADs: (a) The left orbital part of middle frontal gyrus of the AV45 scan. (b) The left angular gyrus of the FDG scan.
major homozygote of rs429358 in ADs were vulnerable to increase beta-amyloid deposition in the left middle frontal gyrus. Similarly, AD patients with the major homozygote of rs429358 were vulnerable to having lower hypometabolism measurement in this ADNI cohort. In a word, these experimental results demonstrated that the proposed MTDFS was very promising in brain imaging genetics.

G. Discussions

We proposed a novel MTDFS method to investigate the nonlinear relationship between brain imaging QTs and SNPs, and simultaneously to identify relevant risk SNPs for brain disorders. To ensure the interpretability, we imposed three sparsity-inducing penalties on the one-to-one layer which was added in front of the multi-tasking DNN. When one applies MTDFS to practical problems, two considerations should be taken. First, due to the multi-tasking modeling, the multiple outputs should be correlated. Otherwise, the single-task based DFS may be a good choice. This indicated that the preselected imaging QTs should have similar changing patterns or genetic basis. Second, the three sparsity-inducing penalties were used here because that SNPs contributed diversely to brain imaging QTs. On this account, they could be replaced by other penalties if one preferred a different sparsity.

To verify the nonlinear and feature selection abilities of MTDFS, we compared it with the two most relevant methods (MTLR and DFS) and used RMSE and CC as metrics. In addition, we conducted correlation analysis between selected SNPs for MTDFS and pre-selected ROI level imaging QTs, as well as population stratification analysis, to explore the genetic variation mechanism of genes. The results showed that our method could identify better subsets of imaging QTs and genetic variations, thereby enhancing the interpretability in brain imaging genetics. Although MTDFS could somewhat improve the interpretability issue of DNN models, it was still insufficient to address the interpretability issue such as how one important loci influence the imaging QTs. Specifically, using MTDFS, we could know which SNPs were nonlinearly related to imaging QTs, but effect sizes of these identified SNPs remained unclear which was an urgent need for the nonlinearly modeling.

IV. CONCLUSION

Extracting the relationship between brain neuroimaging data and genetic data, as well as select relevant genetic factors, is important for brain imaging genetics. The linear model has been extensively studied but is limited since the human genome could nonlinearly affect the brain structure and function. To overcome this drawback, we designed a multi-task deep feature selection method. MTDFS conducted nonlinear relationship extraction and feature selection simultaneously. We introduced a hybrid penalty to select features at the element-sparsity, individual-sparsity, and group-sparsity levels. Results on real neuroimaging genetic data showed that MTDFS was the most

| f11norm | f21norm | G21norm | Task 1 | Task 2 | Task 3 | Task 4 | Task 5 | Task 6 | Task 7 | Task 8 | Task 9 | Task 10 |
|---------|---------|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| ✓       | ✓       | ✓       | 0.6668 | 0.6777 | 0.7226 | 0.7099 | 0.6907 | 0.6913 | 0.7042 | 0.7064 | 0.6957 | 0.6807 |
| ✓       | ✓       | ✓       | (0.016) | (0.016) | (0.016) | (0.014) | (0.013) | (0.009) | (0.011) | (0.013) | (0.019) | (0.018) |
| ✓       | ✓       | ✓       | 0.4043 | 0.4625 | 0.4500 | 0.4605 | 0.4306 | 0.4348 | 0.4472 | 0.4421 | 0.4118 | 0.3953 |
| ✓       | ✓       | ✓       | (0.075) | (0.071) | (0.072) | (0.070) | (0.069) | (0.068) | (0.063) | (0.064) | (0.060) | (0.059) |
| ✓       | ✓       | ✓       | 0.1794 | 0.1673 | 0.2176 | 0.2189 | 0.1914 | 0.1899 | 0.1828 | 0.1943 | 0.1702 | 0.1696 |
| ✓       | ✓       | ✓       | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) | (0.004) | (0.003) | (0.003) |
| ✓       | ✓       | ✓       | 0.4602 | 0.4792 | 0.5321 | 0.5461 | 0.5218 | 0.5285 | 0.5149 | 0.5061 | 0.4714 | 0.4371 |
| ✓       | ✓       | ✓       | (0.023) | (0.029) | (0.021) | (0.021) | (0.021) | (0.021) | (0.021) | (0.021) | (0.021) | (0.021) |
| ✓       | ✓       | ✓       | 0.3964 | 0.3983 | 0.4323 | 0.4354 | 0.3969 | 0.3968 | 0.4247 | 0.3907 | 0.3678 | 0.3570 |
| ✓       | ✓       | ✓       | (0.016) | (0.021) | (0.021) | (0.019) | (0.019) | (0.017) | (0.021) | (0.013) | (0.019) | (0.018) |

The best values in the testing set were shown in bold.
powerful approach among multi-task linear regression and single-task deep feature selection. In the future, we intend to apply the convolution network into our model since it could better identify LD for SNPs, which has the potential to better understand human brain.

ACKNOWLEDGMENT

The first two authors contribute equally to this work. Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904 and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

REFERENCES

[1] S. G. Potkin et al., “Genome-wide strategies for discovering genetic influences on cognition and cognitive disorders: Methodological considerations,” Cogn. Neuropsychiatry, vol. 14, no. 4/5, pp. 391–418, 2009.
[2] L. Shen and P. M. Thompson, “Brain imaging genomics: Integrated analysis and machine learning,” in Proc. IEEE, vol. 108, no. 1, pp. 125–162, Jan. 2020.
[3] S. Ghosal et al., “A generative-discriminative framework that integrates imaging, genetic, and diagnosis into coupled low dimensional space,” NeuroImage, vol. 238, 2021, Art. no. 118200.
[4] N. K. Batmanghelich, A. Dalca, G. Quon, M. Sabuncu, and P. Golland, “Probabilistic modeling of imaging, genetics and diagnosis,” IEEE Trans. Med. Imag., vol. 35, no. 7, pp. 1765–1779, Jul. 2016.
[5] G. D. Pearson, J. Liu, and V. D. Calhoun, “An introductory review of parallel independent component analysis (PICA) and a guide to applying PICA to genetic data and imaging phenotypes to identify disease-associated biological pathways and systems in common complex disorders,” Front. Genet., vol. 6, 2015, Art. no. 276.
[6] A. Arslan, “Imaging genetics of schizophrenia in the post-GWAS era,” Prog. Neuro Psychopharmacol. Biol. Psychiatry, vol. 80, pp. 155–165, 2018.
[7] S. Ghosal et al., “A biologically interpretable graph convolutional network to link genetic risk pathways and imaging phenotypes of disease,” in Proc. Int. Conf. Learn. Representations, 2022, pp. 1–12.
[8] L. Shen et al., “Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: A study of the ADNI cohort,” NeuroImage, vol. 53, no. 3, pp. 1051–1063, 2010.
[9] H. Wang et al., “Identifying quantitative trait loci via group-sparse multitask regression and feature selection: An imaging genetics study of the adni cohort,” Bioinformatics, vol. 28, no. 2, pp. 229–237, 2012.
[10] J. Yan et al., “Transcriptome-guided amyloid imaging genetic analysis via a novel structured sparse learning algorithm,” Bioinformatics, vol. 30, no. 17, pp. i564–i571, 2014.
[11] L. Du et al., “Structured sparse canonical correlation analysis for brain imaging genetics: A novel improved GraphNet method,” Bioinformatics, vol. 32, no. 10, pp. 1544–1551, 2016.
[12] X. Hao et al., “Identification of associations between genotypes and longitudinal phenotypes via temporally-constrained group sparse canonical correlation analysis,” Bioinformatics, vol. 33, no. 14, pp. i341–i349, 2017.
[13] W. Hu et al., “Adaptive sparse multiple canonical correlation analysis with application to imaging (EPI) genetics study of schizophrenia,” IEEE Trans. Biomed. Eng., vol. 65, no. 2, pp. 390–399, Feb. 2017.
[14] L. Du et al., “A novel SCCA approach via truncated $\ell_0$-norm and truncated group lasso for brain imaging genetics,” Bioinformatics, vol. 34, no. 2, pp. 278–285, 2018.
[15] L. Du et al., “Identifying progressive imaging genetic patterns via multi-task sparse canonical correlation analysis: A longitudinal study of the ADNI cohort,” Bioinformatics, vol. 35, no. 14, pp. 1474–1483, 2019.
[16] L. Du et al., “Detecting genetic associations with brain imaging phenotypes in Alzheimer’s disease via a novel structured SCCA approach,” Med. Image Anal., vol. 61, 2020, Art. no. 101656.
[17] L. Du et al., “Associating multi-modal brain imaging phenotypes and genetic risk factors via a dirty multi-task learning method,” IEEE Trans. Med. Imag., vol. 39, no. 11, pp. 3416–3428, Nov. 2020.
[18] L. Du, J. Zhang, F. Liu, H. Wang, L. Guo, and J. Han, “Identifying associations among genomic, proteomic and imaging biomarkers via adaptive sparse multi-view canonical correlation analysis,” Med. Image Anal., vol. 70, 2021, Art. no. 102003.
[19] K. L. Grasby et al., “The genetic architecture of the human cerebral cortex,” Science, vol. 367, no. 6484, 2020, Art. no. eaay6009.
[20] D. P. Hibar et al., “Common genetic variants influence human subcortical brain structures,” Nature, vol. 520, no. 7546, pp. 224–229, 2015.
[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in Proc. Adv. Neural Inf. Process. Syst., 2012, pp. 845–852.
[22] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[23] Y. Li, C.-Y. Chen, and W. W. Wasserman, “Deep feature selection: Theory and application to identify enhancers and promoters,” in Proc. Int. Conf. Comput. Biol., 2015, pp. 205–217.
[24] J. Liu, S. J., and J. Ye, “Multi-task feature learning via efficient 1,1-norm regularization,” in Proc. Conf. Uncertainty Artif. Intell., 2009, pp. 339–348.
[25] W. J. Jagust et al., “The Alzheimer’s disease neuroimaging initiative positron emission tomography core,” Alzheimer’s Dement., vol. 6, no. 3, pp. 221–229, 2010.
[26] J. Jing et al., “Correlation between brain 18F-AV45 and 18F-FDG PET distribution characteristics and cognitive function in patients with mild and moderate Alzheimer’s disease,” J. Alzheimer’s Dis., vol. 79, pp. 1–9, 2020.
[27] T. Thientunyakit et al., “Relationships between amyloid levels, glucose metabolism, morphologic changes in the brain and clinical status of patients with Alzheimers disease,” Ann. Nucl. Med., vol. 34, no. 5, pp. 337–348, 2020.
[28] C. Murphy, “Olfactory and other sensory impairments in Alzheimer disease,” Nature Rev. Neurosci., vol. 15, no. 1, pp. 11–24, 2014.
[29] D. Lin, V. D. Calhoun, and Y.-P. Wang, “Correspondence between fMRI and SNP data by group sparse canonical correlation analysis,” Med. Image Anal., vol. 18, no. 6, pp. 891–902, 2014.
[30] A. Andrew, R. Arora, J. Bilmes, and K. Livescu, “Deep canonical correlation analysis,” in Proc. 30th Int. Conf. Mach. Learn., ser. Proceedings of Machine Learning Research, 2015, pp. 1247–1255.
[31] W. Wang, R. Arora, K. Livescu, and J. Bilmes, “On deep multi-view representation learning,” in Proc. 32nd Int. Conf. Mach. Learn., ser. Proceedings of Machine Learning Research, 2015, pp. 1083–1092.
[32] M. Ghanbari et al., “Genome-wide identification of microRNA-related variants associated with risk of Alzheimer’s disease,” Sci. Rep., vol. 6, no. 1, pp. 1–9, 2016.
[33] A. I. Yashin et al., “Hidden heterogeneity in Alzheimer’s disease: Insights from genetic association studies and other analyses,” Exp. Gerontol., vol. 107, pp. 148–160, 2018.
[34] A. M. Kulminski et al., “Genetic and regulatory architecture of Alzheimer’s disease in the APOE region,” Alzheimer’s Dementia: Diagnosis Assessment Dis. Monit., vol. 12, no. 1, 2020, Art. no. e12008.
[35] L. Du et al., “Identifying diagnosis-specific genotype–phenotype associations via joint multitask sparse canonical correlation analysis and classification,” Bioinformatics, vol. 36, no. Supplement_1, pp. i371–i379, 2020.
[36] H. Xu et al., “High-throughput and efficient multilocus genome-wide association study on longitudinal outcomes,” Bioinformatics, vol. 36, no. 10, pp. 3004–3010, 2020.
[37] D. Clark et al., “Apolipoprotein-E gene variants associated with cardiovascular risk factors in antipsychotic recipients,” Eur. Psychiatry, vol. 24, no. 7, pp. 456–463, 2009.
[38] A. Zhang, Q. Zhao, D. Xu, and S. Jiang, “Brain APOE expression quantitative trait loci-based association study identified one susceptibility locus for Alzheimer’s disease by interacting with APOE,” Sci. Rep., vol. 8, no. 1, pp. 1–5, 2018.
[39] X. Zhou et al., “Identification of genetic risk factors in the Chinese population implicates a role of immune system in Alzheimer’s disease pathogenesis,” in Proc. Nat. Acad. Sci., vol. 115, no. 8, pp. 1697–1706, 2018.
[40] X. Zhou et al., “Non-coding variability at the APOE locus contributes to the Alzheimer’s risk,” Nature Commun., vol. 10, no. 1, pp. 1–16, 2019.
[41] Y. Guo et al., “Genome-wide association study of hippocampal atrophy rate in non-demented elders,” Aging (Albany NY), vol. 11, no. 22, 2019, Art. no. 10468.
[42] T. Horwitz, K. Lam, Y. Chen, Y. Xia, and C. Liu, “A decade in psychiatric GWAS research,” Mol. Psychiatry, vol. 24, no. 3, pp. 378–389, 2019.

Chenglin Yu received the BS degree from the school of microelectronics and control engineering, Changzhou University, Changzhou, China, in 2020. He is currently working toward the graduation degree with the School of Automation, Northwestern Polytechnical University, Xi’an, China. His research interests are in Brain Imaging Genetics, machine learning and deep learning.

Shu Zhang received the PhD degree in computer science from the University of Georgia, USA, in 2018. He is currently a professor with the school of computer science, Northwestern Polytechnical University, Xian, China. His research interests include biomedical image analysis, brain image analysis, deep learning, machine learning algorithms, and artificial intelligence.

Muheng Shang received the BS degree from the school of artificial intelligence, Hebei University of Technology, in 2021. He is currently working toward the PhD degree with the School of Automation, Northwestern Polytechnical University, Xi’an, China. His research interests include brain imaging genetics, machine learning, pattern recognition, and Big Data mining.

Lei Guo received the BS, MS, and PhD degrees in 1982, 1986, and 1993, respectively. He is currently a professor of pattern recognition with Northwestern Polytechnical University, Xi’an, China. His research interests include computer vision, image processing, image segmentation, object detection and tracking.

Junwei Han (Fellow, IEEE) received the PhD degree in pattern recognition and intelligent systems from the School of Automation, Northwestern Polytechnical University, Xi’an, China, in 2003. He is currently a professor in School of Automation, Northwestern Polytechnical University. His research interests include computer vision and multi-media processing.

Lei Du (Member, IEEE) received the PhD degree in computer science from School of the Electronic and Information Engineering, Xi’an Jiaotong University, Xi’an, China, in 2013. Currently, he is an associate professor in School of Automation, Northwestern Polytechnical University, Xi’an, China. His research interests include brain imaging genetics, bioinformatics, machine learning and Big Data mining.