2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Latent Space Learning and Feature Learning using Multi-template for Multi-classification of Alzheimer's Disease

Zihao Chen, Haijun Lei, Zhongwei Huang, and Baiying Lei*, *Member, IEEE*

Abstract— **Alzheimer's disease (AD) is a common brain disease in the elderly that leads to thinking, memory, and behavior disorders. As the population ages, the proportion of AD patients is also increasing. Accordingly, computer-aided diagnosis of AD attracts more and more attention recently. In this paper, we propose a novel model combining latent space learning and feature learning using features extracted from multiple templates for AD multi-classification. Specifically, latent space learning is employed to obtain the inter-relationship between multiple templates, and feature learning is performed to explore the intrinsic relation in feature space. Finally, the most discriminative features are selected to boost the multi-classification performance. Our proposed model uses the data from the Alzheimer's disease neuroimaging initiative dataset. Furthermore, a series of comparative experiments indicate that our proposed model is quite competitive.**

Keywords— **Alzheimer's disease, multi-template features, latent space learning, multi-classification.**

I. Introduction

Alzheimer's disease (AD) is a common neurodegenerative disease in the elderly, which deteriorates over time and progresses slowly [1]. According to the severity of the disease, the progression of AD can be divided into three phases including normal control (NC), mild cognitive impairment (MCI), and AD [2]. Besides, MCI can be further divided into two phases, stable MCI (sMCI) and progressive MCI (pMCI), according to whether the patient can progress to AD in 18 months [2]. It is known that the early diagnosis of AD can help to slow down the progression by monitoring the corresponding stage. Meanwhile, the proportion of AD patients is also increasing as the population ages. Therefore, computer-aided diagnosis (CAD) of AD in the early stage has attracted more and more attention recently.

Many machine learning methods have been proposed for early AD identification [3]. They mainly focus on the analysis of magnetic resonance imaging (MRI) data, which can provide structural information of the human brain. However, conventional methods typically extract features by a single regions-of-interest (ROI) template [3, 4], which is pre-defined to divide the brain into regions. It was proved that using multiple ROI templates to extract multiple sets of features from MRI images can capture richer brain structural information of abnormal brain regions, which is more prospective to compare group differences and reveal disease conditions [5, 6].

However, the features extracted from neuroimaging data using multiple ROI templates are inherently high-dimensional. To reduce the dimensions of feature space and learn more complementary structural information, the inter-relationship between different ROI templates can be considered rather than simply concatenating multiple sets of features together [3]. For example, Chen *et al.* [7] built a multi-task framework to capture the underlying relationship between multi-template features by regarding each set of features as a task. Moreover, as the sample size is limited, the overfitting problem is a serious challenge for AD multi-classification. Accordingly, the intrinsic relationship within feature space should be explored as well. Subspace learning has been employed to discover the inner relation within feature space such as locally preserving projections (LPP) [8] and linear discriminative analysis [9]. Also, sparse learning sets the weight of unimportant features to zero to select informative features by exploring the relation within features space. For example, Nie *et al.* [10] used an $l_{2,1}$-norm to discard the redundant features.

In this article, we propose a novel model integrating latent space learning and feature learning using multi-template features for AD multi-classification. Specifically, we suppose that a latent space exists for multi-template features so that we can project them to this space. In this way, features from different ROI templates can reflect different attribute information. The common latent space will model the inter-relationship between different templates and preserve the complementary structural information. Furthermore, the dimensions of feature space become lower after latent space learning, which is equivalent to a dimensionality reduction operation. After obtaining the common latent space, feature learning that combines LPP and l_{21}-norm is performed to this common space to explore the intrinsic relationship and select the most discriminative features. Specifically, LPP is performed to retain the intrinsic connection within features, and $l_{2,1}$-norm can make the weight matrix sparse to discard unimportant features. After feature learning, the most discriminative features are selected to feed into the classifier for prediction. Finally, data from the Alzheimer's Disease

This work was supported partly by National Natural Science Foundation of China (Nos. 61871274, and 61801305), National Natural Science Foundation of Guangdong Province (Nos.2020A1515010649, and 2019B1515120029), Guangdong Province Key Laboratory of Popular High Performance Computers (No. 2017B030314073), Guangdong Laboratory of Artificial-Intelligence and Cyber-Economics (SZ), Shenzhen Key Basic Research Project (Nos. JCYJ20190808165209410, JCYJ20180507184647636, GJHZ20190822095414576, GJHZ20180418190529516, and JSGG20180507183215520), Special Project in Key Areas of Ordinary Universities of Guangdong Province (No. 2019KZDZX1015).

Z. Chen, H. Lei and Z. Huang are with School of Computer and Software Engineering, Shenzhen University, Guangdong Province Key Laboratory of Popular High-performance Computers, Shenzhen, China.

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

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

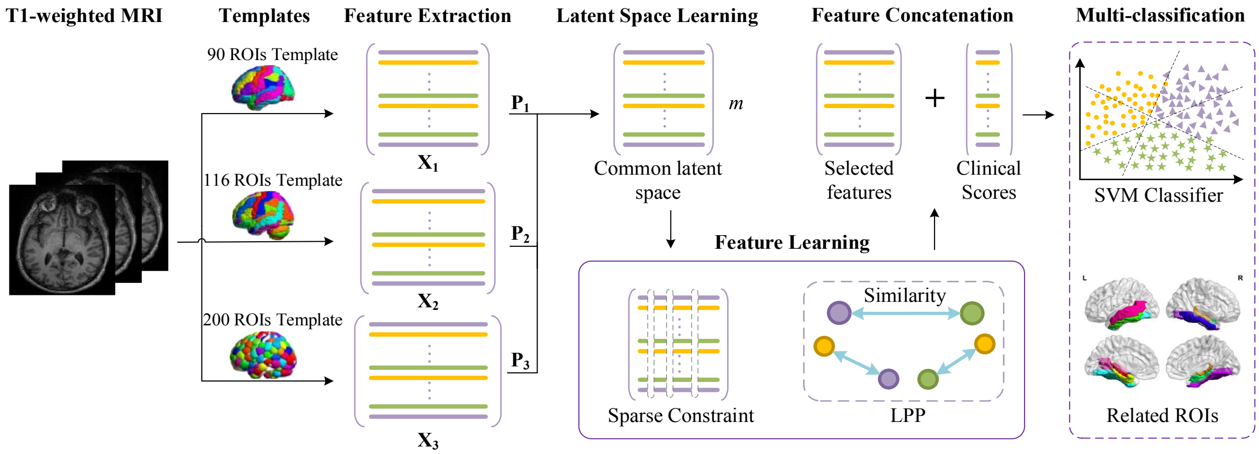


Fig. 1. Overview of our proposed model. We assume that K templates exist. $\mathbf{X}_k \in \mathbb{R}^{n\times d}$ represents the features extracted from k-th template. $\mathbf{P}_k \in \mathbb{R}^{d\times m}$ represents the project matrix corresponding to $\mathbf{X}_k$.

Neuroimaging Initiative (ADNI) database is used to evaluate our proposed model, and the results of comparison experiments prove its competitiveness and superiority.

II. Method

The target of our proposed model is to select the most informative features from high-dimensional multi- template features, thereby improving the multi-classification performance. First, we use multiple ROI templates to extract multi-template features from the original MRI data. Then, the proposed method is used to identify the most informative features from multi-template features. Finally, these selected features concatenated with clinical scores of patients are transferred into the support vector machine (SVM) classifier to classify the subjects into different groups. The overview of our proposed model is displayed in Fig. 1.

A. Proposed Method

We assume that K ROI templates are used in our model. The features extracted from k-th template are referred as $\mathbf{X}_k \in \mathbb{R}^{n\times d}$ and the corresponding label matrix is denoted as $\mathbf{Y} \in \mathbb{R}^{n\times c}$, where d, n, and c indicate the number of feature dimensions, samples, and classes, respectively. We suppose a latent space exists for multi-template features so that these features can be projected into the space. Accordingly, a project matrix $\mathbf{P}_k \in \mathbb{R}^{d\times m}$ is defined to project $\mathbf{X}_k$ to the common latent space $\mathbf{M} \in \mathbb{R}^{n\times m}$, where m represents the dimensions of the latent space. We formulate the following formula to obtain the project matrices.

$$\min_{\mathbf{P}_k,\mathbf{M}} \sum_{k=1}^{K}\|\mathbf{X}_k\mathbf{P}_k-\mathbf{M}\|_{\mathrm{F}}^2+\sum_{k=1}^{K}\|\mathbf{P}_k\|_{2,1}, \tag{1}$$

where $\|\mathbf{A}\|_F=\sqrt{\sum_i\|\mathbf{A}^i\|_2^2}$ is denoted as the Frobenius norm of the matrix $\mathbf{A}$. Besides, the $l_{2,1}$-norm, $\|\mathbf{A}\|_{2,1}=\sum_i\|\mathbf{A}\|_2=\sum_i\sqrt{\sum_j a_{i,j}^2}$, is performed on the project matrices to pick out the most informative features of each template. Considering the inter-relationship between multi-template features, we use different project matrices for each template to obtain the common latent space. Also, after projecting to the latent space, the dimensions of feature space are reduced and the unimportant features in each set of features are discarded, which can alleviate the overfitting issue and preserve the complementary structural information.

After obtaining the common latent space, it can be used as feature space to fit the target matrix by linear regression model $\mathbf{Y}=\mathbf{MW}$, where $\mathbf{W}\in\mathbb{R}^{m\times c}$ is the weight coefficient matrix. Then we can obtain $\mathbf{W}$ by minimizing the formula as follow.

$$\min_{\mathbf{W}}\|\mathbf{Y}-\mathbf{MW}\|_F^2. \tag{2}$$

However, Eq. (2) is a simple linear regression model without any feature selection methods, which will cause poor classification performance. Although the overfitting issue is alleviated using latent space learning, there are still redundant features in the latent common space. Hence, we further explore the intrinsic relationship within the latent space to select the most discriminative features. First, we apply the $l_{2,1}$-norm on the weight coefficient matrix to make the latent space sparse. The $l_{2,1}$-norm first sums the l_2-norm of each row of the weight matrix and then performs l_1-norm on it. The l_1-norm is powerful to make a matrix sparse, and the l_2-norm can prevent overfitting and improve the generalization ability of the model. Thus, the $l_{2,1}$-norm can filter the uninformative features by setting the rows of $\mathbf{W}$ to be zero. Then the objective function is formulated as

$$\min_{\mathbf{W}}\|\mathbf{Y}-\mathbf{MW}\|_F^2+\lambda_1\|\mathbf{W}\|_{2,1}, \tag{3}$$

where λ_1 is the hyper-parameter controlling the regularization term on $\mathbf{W}$. Additionally, we can further investigate the intrinsic relationship within latent space by LPP, which is employed to preserve the local relationship within feature space. In LPP, it first constructs a neighborhood graph for data using k-nearest neighbor and then uses a heat kernel function to compute the similarity matrix $\mathbf{S}\in\mathbb{R}^{m\times m}$. Coupling the $l_{2,1}$-norm and the LPP, we obtain the objective function as

$$\min_{\mathbf{W}}\|\mathbf{Y}-\mathbf{MW}\|_F^2+\lambda_1\|\mathbf{W}\|_{2,1}+$$
$$\lambda_2 tr\left(\sum_{i,j}\left(\mathbf{W}^T\mathbf{m}_i-\mathbf{W}^T\mathbf{m}_j\right)^2 s_{i,j}\right), \tag{4}$$

where λ_1 and λ_2 are penalty factors controlling the $l_{2,1}$-norm and the LPP.

To fully explore the inter-relationship between different templates and the intrinsic relationship within feature space, we integrate latent space learning and feature learning into our proposed model to select the most discriminative features from multi-template features. By combining Eq. (1) and Eq. (4), we obtain the final objective function as follow

$$\min_{\mathbf{W},\mathbf{M},\mathbf{P}_k} \|\mathbf{Y}-\mathbf{MW}\|_F^2 + \lambda_1 tr\left(\sum_{i,j}\left(\mathbf{W}^{\boldsymbol{T}}\mathbf{m}_i - \mathbf{W}^{\boldsymbol{T}}\mathbf{m}_j\right)^{\mathbf{2}} s_{i,j}\right) + \lambda_2\|\mathbf{W}\|_{2,1} + \lambda_3 \sum_{k=1}^{K}\|\mathbf{X}_k\mathbf{P}_k - \mathbf{M}\|_{\mathrm{F}}^2 + \lambda_4 \sum_{k=1}^{K}\|\mathbf{P}_k\|_{2,1}. \tag{5}$$

After selecting the most relevant features by Eq. (5), the selected features concatenated with the clinical scores are fed into the SVM classifier to identify multi-stages of AD.

D. Optimization

In the final objective function, three variables need to be optimized. Therefore, the alternating update method can be used to effectively converge the objective function. We iterate the following three steps to solve Eq. (5): (1) fix $\mathbf{M}$ and $\mathbf{P}_k$ to update $\mathbf{W}$; (2) fix $\mathbf{M}$ and $\mathbf{W}$ to update $\mathbf{P}_k$; (3) fix $\mathbf{P}_k$ and $\mathbf{W}$ to update $\mathbf{M}$.

With fixing $\mathbf{M}$ and $\mathbf{P}_k$, the objective function can be formulated as

$$\min_{\mathbf{W}} \|\mathbf{Y}-\mathbf{MW}\|_F^2 + \lambda_1 tr\left(\sum_{i,j}\left(\mathbf{W}^{\boldsymbol{T}}\mathbf{m}_i - \mathbf{W}^{\boldsymbol{T}}\mathbf{m}_j\right)^{\mathbf{2}} s_{i,j}\right) + \lambda_2\|\mathbf{W}\|_{2,1}. \tag{6}$$

Since Eq. (6) is convex but non-smooth, we cannot directly take the derivative of this formula to obtain the solution of $\mathbf{W}$. According to [8, 11], we employ a new accelerated proximal gradient method to solve it. First, we redefine Eq. (6) as

$$\boldsymbol{K}(\mathbf{W}) = \|\mathbf{Y}-\mathbf{MW}\|_F^2 + \lambda_1 tr\left(\sum_{\boldsymbol{i},\boldsymbol{j}}\left(\mathbf{W}^{\boldsymbol{T}}\mathbf{m}_i - \mathbf{W}^{\boldsymbol{T}}\mathbf{m}_j\right)^{\mathbf{2}} s_{i,j}\right), \tag{7}$$

$$\boldsymbol{\Phi}(\mathbf{W}) = \boldsymbol{K}(\mathbf{W}) + \lambda_2\|\mathbf{W}\|_{2,1}. \tag{8}$$

Then, we can optimize $\mathbf{W}$ with the proximal gradient method to find a closed-form solution of $\mathbf{W}$.

With fixing $\mathbf{M}$ and $\mathbf{W}$, we can formulate the objective function as

$$\min_{\mathbf{P}_k} \lambda_3 \sum_{k=1}^{K}\|\mathbf{X}_k\mathbf{P}_k - \mathbf{M}\|_{\mathrm{F}}^2 + \lambda_4 \sum_{k=1}^{K}\|\mathbf{P}_k\|_{2,1}. \tag{9}$$

We can observe that Eq. (9) is similar to Eq. (6). Consequently, the same method can be employed to optimize Eq. (9). Then, we can obtain the solution of $\mathbf{P}_k$.

With fixing $\mathbf{P}_k$ and $\mathbf{W}$, we can formulate the objective function as

$$\min_{\mathbf{W}} \|\mathbf{Y}-\mathbf{MW}\|_F^2 + \lambda_1 tr\left(\sum_{i,j}\left(\mathbf{W}^{\boldsymbol{T}}\mathbf{m}_i - \mathbf{W}^{\boldsymbol{T}}\mathbf{m}_j\right)^{\mathbf{2}} s_{i,j}\right) + \lambda_3 \sum_{k=1}^{K}\|\mathbf{X}_k\mathbf{P}_k - \mathbf{M}\|_{\mathrm{F}}^2. \tag{10}$$

We take the derivative of this formula and make it equal to zero to obtain the solution of $\mathbf{M}$.

Accordingly, we alternately update the values of $\mathbf{W}$, $\mathbf{M}$ and $\mathbf{P}_k$, and the objective function will finally converge to obtain the optimal $\mathbf{W}$, $\mathbf{M}$ and $\mathbf{P}_k$. Then, we can use the optimal $\mathbf{W}$ to select the most discriminative features, which leads to the superior performance of multi-classification.

III. Experiments

A. Experiment Setting

According to different phases of AD, we set two multi-classification tasks in our experiments, including NC vs. MCI vs. AD denoted as AD3, and NC vs. sMCI vs. pMCI vs. AD denoted as AD4. In our objective function, we have four penalty factors, which are λ_1, λ_2, λ_3 and λ_4. We find the optimal value using the grid searching strategy. Different metrics, including accuracy (ACC), sensitivity (SEN), precision (PREC), and F1-score, are used to evaluate our multi- classification performance. To verify the effectiveness of our method, we use a 10-fold cross-validation strategy in our experiments.

B. Data and Preprocessing

In this article, we use MRI data from the ADNI dataset (https://www.loni.usc.edu/). 814 subjects including 220 NC, 402 MCI, and 192 AD subjects are collected for the AD3 task. Furthermore, 402 MCI subjects can be divided into 256 sMCI subjects and 146 pMCI subjects for the AD4 task. Also, we collect mini-mental state examination (MMSE) scores as the clinical scores from ADNI as features. The experimental procedures involving human subjects described in this paper were approved by the Institutional Review Board.

After collecting original MRI data from the ADNI dataset, the anterior commissure-posterior commissure correction and skull-stripping [12] are first performed on each image for the next precise segmentation. Then, we use the statistical parametric mapping toolbox [13] to segment these images into the cerebrospinal fluid (CSF), gray matter (GM), and white matter (WM), which are remarkable tissues in the brain images and widely used in other researches. Finally, we use three ROI templates, including automatic anatomical labeling (AAL) atlases [14] and Craddock's spatially constrained spectral clustering atlas [15], to divide 90, 116, and 200 ROIs for CSF, GM, and WM, respectively, and then the mean tissue density value of each ROI can be computed as features.

C. Experimental Results

In this article, we set a series of comparison experiments with several methods to evaluate the superiority of our proposed methods such as Lasso [16], M3T [17], LRpL1 [18], SLRL [19], TMSLRL [7]. These methods are all used for feature selection and use the same SVM classifier with the RBF kernel. The detailed results of comparative experiments are listed in Tables I and II. We can observe that the results of our proposed model can get better results than other competitive models in all evaluation metrics and all classification tasks. Moreover, we add all confusion matrices of 10-fold cross-validation experiments together and plot them in Fig. 2. We can observe that our proposed model is more effective mainly due to the powerful ability of identifying MCI, sMCI, and pMCI.

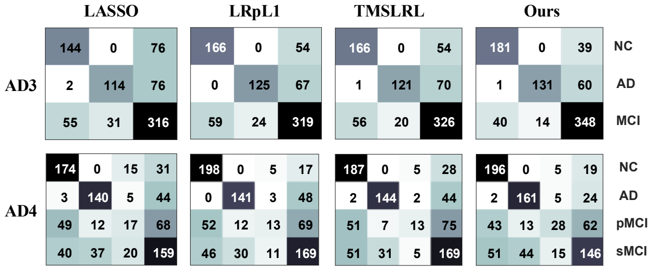
The top ten brain regions related to AD obtained by the optimal weight matrix of our proposed method for two multi-classification tasks are visualized in Fig. 3, which may assist researchers to further study AD in the future. We can see that the brain areas obtained by the two tasks show no obvious difference, which is a reasonable result.

TABLE I. THE RESULTS OF AD3 (MEAN±STANDARD DEVIATION).

Method	ACC	SEN	PREC	F1-score
LASSO	70.50±5.27	67.81±6.25	73.35±5.48	70.38±5.34
M3T	70.88±5.22	68.15±5.53	73.97±5.77	70.84±4.91
SLRL	75.30±4.38	73.89±4.74	78.09±5.44	75.83±4.15
LRpL1	74.93±5.6	73.33±5.85	77.21±5.75	75.17±5.44
TMSLRL	75.30±5.64	73.21±6.32	78.29±5.21	75.6±5.44
Ours	**81.07±4.71**	**79.01±6.04**	**84.11±3.67**	**81.41±4.49**

TABLE II. THE RESULTS OF AD4 (MEAN±STANDARD DEVIATION).

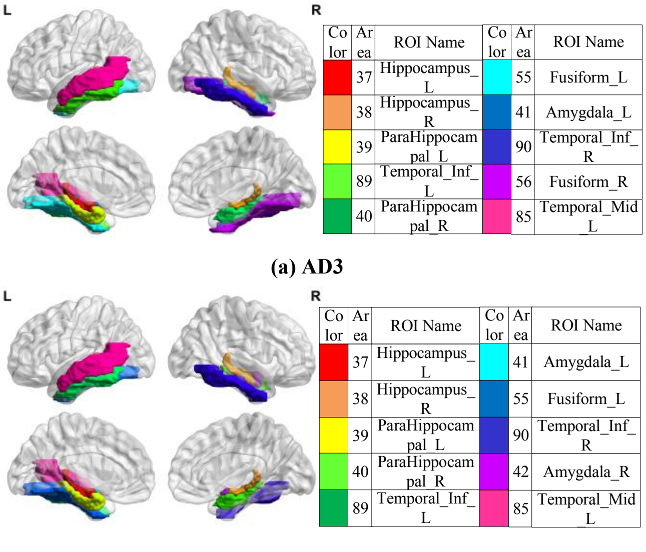
Method	ACC	SEN	PREC	F1-score
LASSO	60.19±2.55	56.50±2.99	55.36±5.57	55.82±3.80
M3T	59.58±2.91	55.42±3.31	55.41±8.36	55.25±5.31
SLRL	59.82±3.15	56.05±3.32	57.8±9.94	56.67±6.0
LRpL1	63.01±5.26	58.69±5.77	56.06±10.93	57.08±7.64
TMSLRL	63.99±4.22	59.64±3.61	60.42±9.83	59.8±6.08
Ours	**65.23±4.04**	**62.29±3.49**	**61.52±11.74**	**61.49±7.07**



	LASSO			LRpL1			TMSLRL			Ours			
AD3	144	0	76	166	0	54	166	0	54	181	0	39	NC
	2	114	76	0	125	67	1	121	70	1	131	60	AD
	55	31	316	59	24	319	56	20	326	40	14	348	MCI

	LASSO				LRpL1				TMSLRL				Ours				
AD4	174	0	15	31	198	0	5	17	187	0	5	28	196	0	5	19	NC
	3	140	5	44	0	141	3	48	2	144	2	44	2	161	5	24	AD
	49	12	17	68	52	12	13	69	51	7	13	75	43	13	28	62	pMCI
	40	37	20	159	46	30	11	169	51	31	5	169	51	44	15	146	sMCI

Fig. 2. The confusion matrices of comparative experiments.



Co lor	Ar ea	ROI Name	Co lor	Ar ea	ROI Name
🟥	37	Hippocampus_L	🟦	55	Fusiform_L
🟧	38	Hippocampus_R	🟦	41	Amygdala_L
🟨	39	ParaHippocampal_L	🟦	90	Temporal_Inf_R
🟩	89	Temporal_Inf_L	🟪	56	Fusiform_R
🟩	40	ParaHippocampal_R	🟥	85	Temporal_Mid_L

(a) AD3

Co lor	Ar ea	ROI Name	Co lor	Ar ea	ROI Name
🟥	37	Hippocampus_L	🟦	41	Amygdala_L
🟧	38	Hippocampus_R	🟦	55	Fusiform_L
🟨	39	ParaHippocampal_L	🟦	90	Temporal_Inf_R
🟩	40	ParaHippocampal_R	🟪	42	Amygdala_R
🟩	89	Temporal_Inf_L	🟥	85	Temporal_Mid_L

(b) AD4

Fig. 3. The top ten ROIs related to AD obtained by the optimal weight matrix of our proposed method.

IV. CONCLUSION

In this paper, we propose a multi-classification model combining latent space learning and feature learning to select informative features from multi-template features. With selected features, the SVM classifier is used to conduct multi-classification of AD. Specifically, we first extract the interrelationship between different templates to the common latent space. Feature learning is performed on the latent space to explore the intrinsic relation to discover the most discriminative features. Finally, a series of comparative experiments illustrate that our proposed model achieves the best performance compared to competing models using the data collected from the ADNI dataset.

REFERENCES

[1] M. Liu, J. Zhang, P.-T. Yap, and D. Shen, "View-aligned hypergraph learning for Alzheimer's disease diagnosis with incomplete multi-modality data," *Med. Image Anal.,* vol. 36, pp. 123-134, 2017.

[2] "2020 Alzheimer's disease facts and figures," vol. 16, no. 3, pp. 391-460, 2020.

[3] B. Lei *et al.*, "Adaptive sparse learning using multi-template for neurodegenerative disease diagnosis," *Med. Image Anal.,* vol. 61, p. 101632, 2020.

[4] T. Zhou, M. Liu, K. H. Thung, and D. Shen, "Latent Representation Learning for Alzheimer's Disease Diagnosis With Incomplete Multi-Modality Neuroimaging and Genetic Data," *IEEE Trans. Med. Imaging,* vol. 38, no. 10, pp. 2411-2422, 2019.

[5] Y. Jin *et al.*, "Identification of infants at high-risk for autism spectrum disorder using multiparameter multiscale white matter connectivity networks," *Hum. Brain Mapp.,* vol. 36, no. 12, pp. 4880-96, 2015.

[6] M. Liu, D. Zhang, E. Adeli, and D. Shen, "Inherent Structure-Based Multiview Learning With Multitemplate Feature Representation for Alzheimer's Disease Diagnosis," *IEEE Trans. Biomed. Eng.,* vol. 63, no. 7, pp. 1473-82, 2016.

[7] Z. Chen *et al.*, "Template-Oriented Multi-task Sparse Low-Rank Learning for Parkinson's Diseases Diagnosis," Cham, 2020, pp. 178-187.

[8] X. Zhu, H. I. Suk, S. W. Lee, and D. Shen, "Subspace Regularized Sparse Multitask Learning for Multiclass Neurodegenerative Disease Identification," *IEEE Trans. Biomed. Eng.,* vol. 63, no. 3, pp. 607-618, 2016.

[9] G. C. Lin, W. J. Wang, C. M. Wang, and S. Y. Sun, "Automated Classification of Multi-Spectral MR Images Using Linear Discriminant Analysis," *Comput. Med. Imag. Grap.,* vol. 34, no. 4, pp. 251-268, 2010.

[10] F. Nie, H. Huang, X. Cai, and C. Ding, "Efficient and robust feature selection via joint $\ell_{2,1}$-norms minimization," *in Proceedings of the 23rd International Conference on Neural Information Processing Systems - vol. 2*, New York, 2010, pp. 1813-1821.

[11] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*. Spring, 2004.

[12] S. A. Sadananthan, W. Zheng, M. W. L. Chee, and V. Zagorodnov, "Skull stripping using graph cuts," *NeuroImage,* vol. 49, no. 1, pp. 225-239, 2010.

[13] W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E. Nichols, *Statistical parametric mapping: the analysis of functional brain images,* 2006.

[14] N. Tzourio-Mazoyer *et al.*, "Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain," *NeuroImage,* vol. 15, no. 1, pp. 273-289, 2002.

[15] R. C. Craddock, G. A. James, P. E. Holtzheimer III, X. P. Hu, and H. S. Mayberg, "A whole brain fMRI atlas generated via spatially constrained spectral clustering," *Hum. Brain Mapp.,* vol. 33, no. 8, pp. 1914-1928, 2012.

[16] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *J. R. Stat. Soc. B,* vol. 58, no. 1, pp. 267-288, 1996.

[17] 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, no. 2, pp. 895-907, 2012.

[18] E. Adeli, X. Li, D. Kwon, Y. Zhang, and K. Pohl, "Logistic Regression Confined by Cardinality-Constrained Sample and Feature Selection," *IEEE Trans. Pattern Anal. and Machine Intell.,* vol. 42, pp. 1713-1728, 2019.

[19] H. Lei, Y. Zhao, Z. Huang, F. Zhou, L. Huang, and B. Lei, "Multi-classification of Parkinson's Disease via Sparse Low-Rank Learning," in *2018 24th International Conference on Pattern Recognition (ICPR)*, 2018, pp. 3268-3272.