CNN-Based Early Mild Cognitive Impairment Diagnosis Using Structure MR Images

Jingwan Jiang,Li Kang,Jianjun Huang and Tijiang Zhang

***Abstract*—Mild cognitive impairment (MCI) is an early sign of Alzheimer's disease (AD) which is the fourth leading disease mostly found in the aged population. In the recent years, extensive attention has been drawn towards AD due to its increasing incidence. Early intervention of MCI will possibly delay the progress towards AD, and this makes it very important to diagnose early MCI(EMCI). However, most studies of discriminating EMCI from cognitively normal control (NC) based on traditional machine learning algorithm suffer from some serious limitations, such as lower accuracy with the subtle difference between EMCI and NC. In this paper, a CNN-based diagnosis approach using structure MRI images was presented for improving classification performance. Specifically, we develop a VGG16-based Convolutional Neural Network (CNN) for exploiting deeply embedded diagnosis features with transfer learning technique; then a feature selection strategy is performed to eliminate redundant features. A Support Vector Machine (SVM) is further employed to distinguish EMCI from NC. Experiments were performed on the publicly available ADNI dataset with a total of 120 subjects. The classification results demonstrate the superior performance of the proposed method with accuracy of 89.4% for EMCI versus NC. Additionally, the transfer learning technique replacing feature extraction based on prior knowledge not only realizes deep learning with small dataset, but also reduces experimental time greatly.**

***Index Terms*—Early Mild Cognitive Impairment, Convolutional Neural Network, Support Vector Machine, Transfer learning**

I. Introduction

ALZHEIMER'S disease, the most common form of dementia, is an irreversible progressive neurodegenerative disease happening to people over the age of 65. AD patients will suffer from memory loss, which disrupts daily life, such as confusion with time or place. In 2018, the estimated number of AD patients were 5.7 million in American, 0.5 million more than in 2014 [1], [2]. However, if no precautions were taken, with this increasing incidence rate there will be 13.8 million AD patients in 2050 [2]. At present, no effective prevention or treatment has been found. More importantly, it costs a lot of money and manpower to cure AD patients. Consequently, early precise intervention of AD will have great personal and financial benefits, which can delay the progression to AD and save up to $7.9 trillion in medical and care costs [2].

Mild cognitive impairment (MCI) is the transition state between age-related cognitive decline and AD or another dementia [3]. People with MCI have mild but measurable changes in thinking abilities. A systemic review of 32 researches found that an average of 32 percent of individuals with MCI develop into AD within 5 years follow-up, which means people with MCI has high risk of conversion to AD [2]. Therefore early diagnosis of MCI is of vital importance for the early intervention in preclinical state of AD [4] [5], and it has drawn much attention of researchers in the last decades [6]. However, the diagnosis of early MCI (EMCI) is one of the most challenging clinical problems because of subtle difference between EMCI and NC, and till now very less research has been done to detect early MCI comparing to AD. In this paper, we have summarized state-of-the-art methods about EMCI and further studied the classification method of distinguishing EMCI from NC group.

II. Related work

Nowadays, neuroimaging techniques have been widely applied to medical image analysis, such as structure magnetic resonance imaging (sMRI), functional MRI (fMRI), diffusion tensor imaging (DTI) and positron emission tomography (PET). These neuroimaging techniques have also been applied for the diagnosis of EMCI. In the literature review, traditional machine learning approaches are used for classifying EMCI and NC by most researches.

Based on 68 cortical areas of diffusion weighted MR images, Prasad et al. [7] computed a 68×68 connectivity matric and a set of network measures as the input of SVM classifier and acquired a classification accuracy of 59.2% for EMCI versus NC. Eunsong Kang et al. [8] proposed a method of probabilistic source separation where the authors separated the fMRI signals into a true source signal and a noise component by a stochastic variational Bayesian inference. Then the relations of the inferred source signals are taken as the input features of a linear SVM, achieving an accuracy of 74.45%. Rory Raeper et al. [9] constructed a cooperative correlational and discriminative ensemble learning framework using sMRI images where each individual brain was represented by a set of shallow convolutional brain multiplex (SCBM) used to train an ensemble of CCA-SVM and LDA-based classifiers, and an accuracy of 80.95% was reported. Parisa Forouzannezhad et al. [10] combined the features extracted from cortical region and subcortical region of MRI and PET images, neuropsychological test scores, age and education to train a deep neural network for EMCI classification and reported an accuracy of 84%. In addition, the authors trained a SVM classifier utilizing the same data in paper [10] and reported an accuracy of 81.1% [11].

The aforementioned approaches based on traditional machine learning try to find distinguishing features from neuroimaging data through complex feature extraction, then perform classification task. However, the process of the above feature extraction not only needs accurate prior knowledge, but it also takes a lot of time and manpower. In addition, In the above mentioned approaches, the diagnosis accuracy of EMCI is far behind that of accuracies achieved by AD or MCI due to the subtle but complex difference between EMCI and NC. The low accuracy score reveals that it is very challenging to distinguish EMCI from NC just utilizing low-level and coarse-grained features based on prior knowledge [12]–[14]. Recently, researchers have shown an increasing interest in convolutional neural network (CNN) method for classification task [15] [16]. CNN is regarded as state-of-the-art machine learning algorithm in the field of image classification, which can extract low- and high-level features from complex high-dimensional image data in the form of end-to-end. There have been several relevant investigations into the diagnosis of EMCI using the popular CNN method.

Tae-Eui Kam et al. [12] proposed a novel 3DCNN framework using fMRI data to extract deep embedded features from both static and dynamic brain functional networks (BFNs) for EMCI diagnosis and reported an accuracy of 76.07%. Lulu Yue et al. [17] utilized deep CNN approach to extract the most useful features of the gray matter of sMRI, and then these features were used to classify EMCI and AD, LMCI and EMCI. However, the classification task for EMCI and NC did not be performed. Mukul Puranik et al. [18] employed Inception Resnet V2 deep learning model with transfer learning technique to classify NC, EMCI and AD, and obtained an accuracy of 98.41%. However, the input data are the 2D slices of fMRI images, which means the classification task isn't based on subject-level, deviating clinical needs.

The above deep learning approaches did not engage with the problem of binary classification of EMCI versus NC except paper [12]. However, time-consuming, multi-channel and multi-model training in paper [12] did not exchange higher classification accuracy. In order to overcome the limitation of machine learning approaches and effectively solve the challenging binary classification problem for EMCI versus NC, a hybrid diagnosis method based on deep CNN and traditional machine learning techniques is proposed in this paper, which achieves higher accuracy in classifying EMCI and NC features using sMRI images. In addition, transfer learning technique is employed to train CNN, which can alleviate the problems caused by small dataset and reduce a lot of training time at the same time.

III. Material and methods

A. Dataset and Preprocessing

Data used in this study were obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI) project launched in 2003 as a public-private partnership. The goal of ADNI study is to detect AD at the earliest possible stage (EMCI, MCI, LMCI) and support advance in intervention, prevention and treatment through new diagnostic methods.

A total of 120 preprocessed MRI scans in NIfTI file format were downloaded from ADNIGO and ADNI2 database. All subjects from the basebline/screening visit have passed strict inclusion criteria. These 120 scans consists of 70 EMCI subjects and 50 age-matched Normal Control (NC). The corresponding demographic information of all 120 subjects is shown in Table I where Mini Mental State Examination (MMSE) is a mental test quantifying the cognitive function, just as a auxiliary diagnostic index. Lower MMSE indicates poor cognitive ability.

TABLE I
Corresponding statistical information of subjects

	EMCI	**NC**
Number	70	50
Gender(F/M)	27/43	27/23
Age(year)	72.9±8.3	72.5±6.1
MMSE	27.86±1.66	28.93±1.18

The raw T1-weighted structure MR images (sMRI) were acquired by 3-Tesla GE medical systems scanners at multiple sites with rigorous quality control to reduce site effect. The following imaging parameters were used: volume size=256×256×196, voxel size=1.0×11.0×11.2 mm^3, flip angle=11°. More information about the parameters of the images can be searched on the website of ADNI (http://adni.loni.usc.edu/). All obtained data have been preprocessed through a series of standard preprocessing procedures. Using FSL and FreeSurfer software, the raw T1-weighted structure MR images are preprocessed with skull-stripping, intensity normalization and registration with a standard template Colin27 having the same coordinate system as MNI152. Finally, the sMRI image of each subject has a resolution of 110×110×110 voxels.

In order to highlight most distinguishable features enhancing efficiency of classification task, 2D slices data of sMRI were studied. For each sMRI image, 32 slices with indices 37 to 68 were saved in form of JPEG image format. It is worth noting that some brain regions associating with memory, such as hippocampus and callosum, are contained in 32 slices. Then a subject-level dataset for SVM classification was built, which consists of 120 folders each of which includes 32 slices of one subject. In addition, another slice-level dataset was built for training 2DCNN model, which contains 2240 (32×70) EMCI slices and 1600 (32×50) NC slices split into train set and validation set with the ratio of 8:2.

B. Proposed Method

The pipeline of proposed approach for EMCI classification is shown in Fig. 1, where a CNN model of VGG16 [19] is chosen as feature extractor, LASSO algorithm is utilized for feature selection and SVM classifier performs classification task. Firstly, VGG16 network is fine-tuned by slice-level dataset using transfer learning technique through loading pre-trained weight, then the optimal VGG16 model is saved in term of the minimum loss value during training process. Secondly, the features of each slice in subject-level dataset are extracted by VGG16 optimal model whose output is a matrix of 1×256. Each subject has 32 slices; therefore 32 matrixes of 1×256 are concatenated into a total feature matrix as a feature

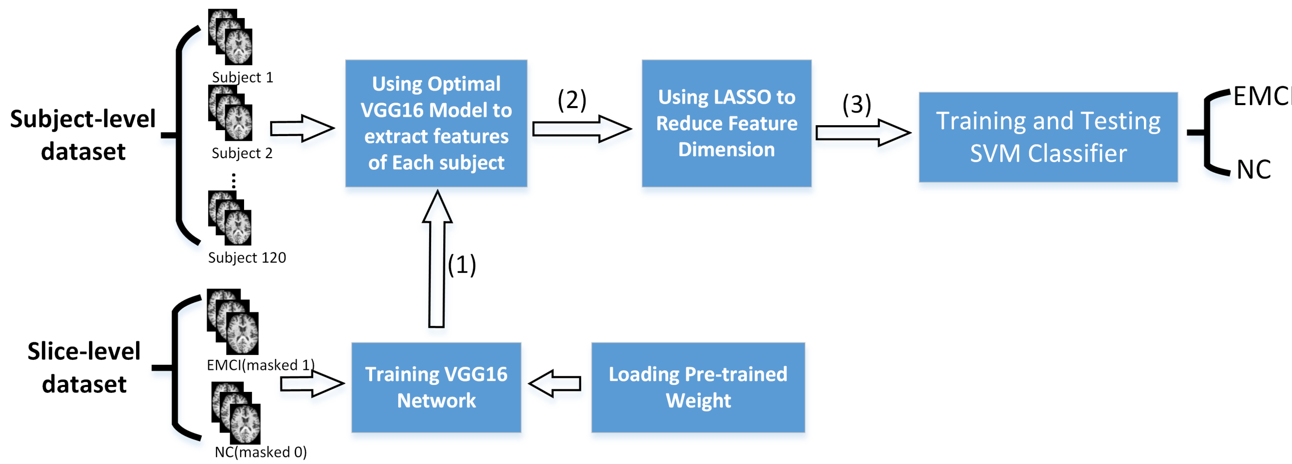


Fig. 1. An illustration of the pipeline of proposed method, where (1) is the optimal VGG16 model which is used as feature extractor in next step, (2) is a matrix with the size of 120×8192(32×256), (3) represents a matrix with the size of 120×*number of selected features of each subject*.

representation of one subject. All feature representations of 70 EMCI and 50 NC subjects are integrated into a matrix dataset, as described in Fig.1 (2). Thirdly, feature selection is performed by LASSO algorithm to reduce the dimension and irrelevant information of the above matrix dataset. Finally, the output of LASSO algorithm is used to train and test SVM classifier for binary classification of EMCI and NC.

1) Convolutional neural network and Transfer learning: Convolutional neural network (CNN) is one of popular deep learning algorithm, which has gotten great success in computer vision and image processing applications in recent years. CNN commonly consists of convolution layer, pooling layer and fully connected layer [17]. In convolution layer, the convolution calculation is performed on an image of size $h \times w$ using a kernel size of k, padding of p and stride of s, then $2^n (n \in Z)$ feature maps with a size of $(\frac{h-k+2p}{s} + 1) \times (\frac{h-k+2p}{s} + 1)$ will be output. The convolution kernels play vital roles, just like feature detector which can learn general and fine-grained features, such as edge, shape and some hidden information [20]. With the similarity features among adjacent regions, the pooling layer can reduce redundant information through acquiring the maximum or mean of a region, and therefore high feature dimensions, vast network parameters and long training time will be reduced drastically. In the fully connected layer, all neurons have full connection to the output of the previous layer. The fully connected layer converges all learned features to classify, finally outputs the classification score and give the actual prediction of input data. Fig. 2 shows the network structure of VGG16 CNN model which consists of five convolution blocks and a fully connected layer containing one flatten layer and two dense layers.

However, the training of VGG16 network in Fig.1 with small dataset will cause overfitting problem in a large probability. Therefore we used dropout and transfer learning technique to alleviate overfitting phenomenon. Dropout technique is to add dropout layers to VGG16 network, which set the output of neurons in the hidden layers into 0 with a probability of r. Transfer learning technique is to utilize the pre-trained weights to initialize own network whose structure is the same as pre-trained model trained by a much larger dataset, thus realizes self-adaption from source to target domain [21] [22]. In this study, we transferred VGG16 pre-trained weights trained by nature image dataset *Imagenet* with 1000 categories into own VGG16 network. Although brain image is very different from nature images, the first few layers in CNN can extract many generic features, such as side, angle, colour et al. According to the difference of category number and image attribute between source domain and target domain, we replaced fully connected layer and freezed the pre-trained weights of the first four convolution blocks of VGG16, then the pre-trained weights in the fifth convolution block and the initial weights of fully connected layer were continually updated during training process, the above courses are also called fine-tuning. The freezed layers will be used to extract generic feature while the fine-tuned layer extract high-level target-specific features.

2) LASSO: With high-dimensional features of each subject, the feature selection algorithm of least absolute shrinkage and selection operator (LASSO) was performed to remove the irrelevant or redundant features so that the feature dimensions would be reduced drastically and overfitting phenomenon can be alleviated effectively. LASSO is performed through minimizing the penalized objective function with L1 regularization, which tends to give zero weight to irrelevant features; therefore the useful discriminative features can be saved [23]. The objective function of LASSO is defined as follows:

$$J(\theta) = \frac{1}{2}\|Y - X^T\theta\|_2^2 + \lambda\|\theta\|_1 \quad (1)$$

where $X = [x_1, x_2, ..., x_N] \in R^{N \times d}$ is a feature matrix. N is the number of subjects and d is the number of features of each subject, thus each x_i represents all features of one subject. $Y = \{y_i | y_i \in \{-1, +1\}\}_{i=1}^N$ is a set of corresponding class labels of subjects. θ represents the regression coefficient and λ is the regularization parameter to balance the complexity of the model.

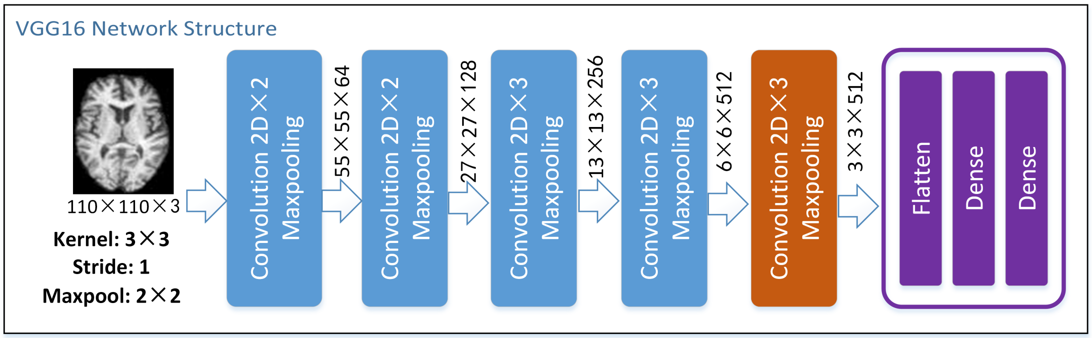


Fig. 2. VGG16 network structure, the gray-scale image in target domain with the size of 110×110×1 will be converted into the format of RGB in source domain with the size of 110×110×3.

3) SVM: With a small and high-dimensional dataset, support vector machine is most appropriate classifier to distinguish EMCI from NC. Given a training set $\{x_k, y_k\}_{k=1}^{N}$ with input data $x_k \in R^n$ and corresponding binary class labels $y_k \in \{-1, +1\}$, the output of primal SVM is presented as follows:

$$y(x) = sign[w^T \varphi(x) + b] \tag{2}$$

Where $\varphi(x)$ is a nonlinear function mapping the input space to higher dimensional feature space, which makes the input data linearly separable in hyperplane. The term b is a bias term. The optimization objective function is defined as follows [24]:

$$\min_{w,b,\xi} J(w, \xi) = \frac{1}{2} w^T w + c \sum_{k=1}^{N} \xi_k \tag{3}$$

subject to:

$$y_k[w^T \varphi(x_k) + b] \geq 1 - \xi_k, k = 1, ..., N, \xi_k \geq 0 \tag{4}$$

ξ_k is slack variable which can allow model to appear misclassification. W is the weight applied for input data x. The positive real constant c is a tuning parameter.

C. Implementation

The training of VGG16 network is implemented based on Keras with a single GPU (i.e. NVIDIA GTX TITAN 12GB). The network is optimized by root mean square propagation (RMSProp) with a learning rate of 10^{-4}. The weight update is performed in mini-batches of 32 samples per batch and stops after 100 epochs.

IV. Experimental and Result

A. Experimental Setting

In this section, several experiments are designed to validate the efficiency of the proposed method in this paper. More specifically, we want to know how the proposed approach improve the classification performance step by step. We first implemented all the steps in Fig.1, then in order to highlight the advantage of transfer learning technique, we also trained VGG16 model from scratch without transferring pre-trained weight. It is worth noting that the training of VGG16 using transfer learning saved half the time comparing with training from scratch. Then, to demonstrate the effectiveness of feature selection algorithm, we also made another contrast test where LASSO algorithm did not perform before training SVM classifier. Finally, we compared the proposed approach with other present methods based on sMRI and EMCI diagnosis. In the following, there are some corresponding experimental setup illustrated below.

As described in section III, the regularization parameter λ of LASSO feature selection algorithm can balance the model complexity. In many trial experiments, we found that the magnitude of λ will influence the classification results to some extent. The variation curve of classification accuracy of EMCI and NC changing with λ is shown in Fig.3, where the range of λ varies from 0 to 8 and the interval is 0.1. However, we did not select the λ corresponding to the highest accuracy because the selected features will lead to overfitting in classification. In order to acquire a relatively robust classification model, a appropriate λ would be selected through repeating experiment with different λ.

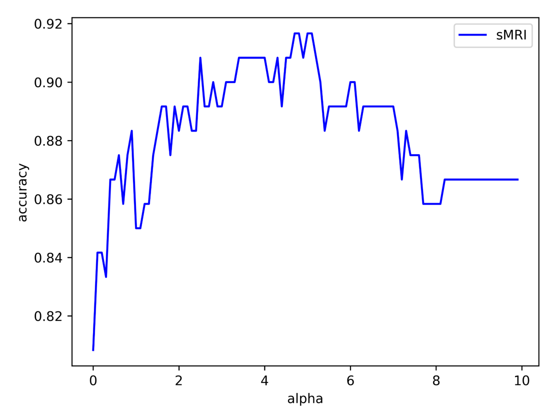


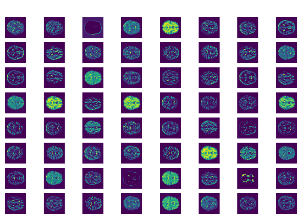
Fig. 3. The variation curve of classification accuracy changing with λ

To alleviate the effect of data abnormity as far as possible, the original data distribution of features was normalized as a normal distribution with unit standard deviation and zero mean before SVM classification. In order to obtain the optimal parameters of SVM classifier, we used *GridSesearchCV* [25] method to generate the training results of each combination of parameters through exhaustive search. The parameters dictionary will be offered by researchers. For example, given a group of parameters $\{C : 10^{-5}, 10^{-4}, ..., 10^{5}\}$ corresponding to linear kernel function, there are 11 times classification task performed with 11 different penalty coefficient C, the best parameter will be chosen through classification results. In order to obtain robust results, we repeated SVM classification task with best parameters 10 times using stratified 10 folds cross-validation strategy and acquired the mean of every metric as the final results. Fig.3 shows the result of one time experiment. In addition, in order to alleviate the problem of data imbalance, class weight was imposed during training.

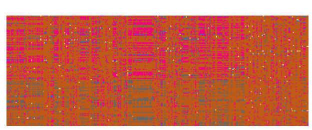
B. Features Extraction and Visualization

CNN is believed to have great ability of extracting class-discriminative features from images, and visualization of the activation map of features is a helpful approach to explore training progress of CNN model. We visualized the output of the first Maxpooling layer of VGG16 model, as shown in Fig.4 (a). Different filter (or convolution kernel) learns different features from various aspects, for example, some filters learn the brain shape, and others learn the interior structure of brain. In Maxpooling layer, after taking the maximum value within a 2×2 region in turn in whole feature map, the size of feature map in previous layer was decreased in half and the details became more obvious. The features will become more and more sparse and localized in deeper layers, at the same time, the embedded high-level abstract features can be learned.

Fig.4 (b) shows the feature matrix output of LASSO algorithm, which was also described in Fig.1. There are apparent distinction between 70 EMCI and 50 NC subjects in Fig.4 (b) where it is clearly seen that the color of the features of NC is more green while that of EMCI is more pink. The well discrimination between EMCI and NC benefited by the ability of feature extraction of VGG16 model to a large extend.



(a)



(b)

Fig. 4. (a) 64 feature activation maps of a sMRI slice with the size of 55×55 pixels, the output of the first Maxpooling layer of VGG16 model. (b) The feature matrix with the size of 120×330 from feature selection.

C. Performance Evaluation

We quantitatively evaluated the classification performance of the proposed method based on accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the receiver operating charateristic curve (AUC). The ROC curves of one 10-folds cross-validation are shown in Fig.5. To our best knowledge, we are the first to combine CNN with traditional machine learning algorithm for EMCI diagnosis. In order to compare the classification performance with other recent methods reasonably, we chose these researches which used sMRI data from ADNI website, as shown in Table II. With the same metrics, it can be clearly seen that the proposed method yielded the best results whatever ACC, SEN or SPE. An mean AUC of 96% reported in Fig.5 also illustrates that the classifier has fairly well classification capacity for positive samples and negative samples.

TABLE II

Classification performance comparison with pervious researches using sMRI for distinguishing EMCI from NC. Where *from scratch* and *no LASSO* respectively represents transfer learning technique and LASSO algorithm did not performed in experiment. Fine-tune & LASSO is the complete process of the proposed method.

Method	ACC(%)	SEN(%)	SPE(%)
Forouzannezhad et al.[10]	61.1	66.5	58.7
Forouzannezhad et al.[11]	73.1	76.8	68.4
Raeper et al.[9]	81.0	83.3	78.57
ours (from scratch)	63.3	72.0	56.3
ours (no LASSO)	80.0	84.0	75.6
ours (fine-tune & LASSO)	**89.4**	**92.9**	**89.3**

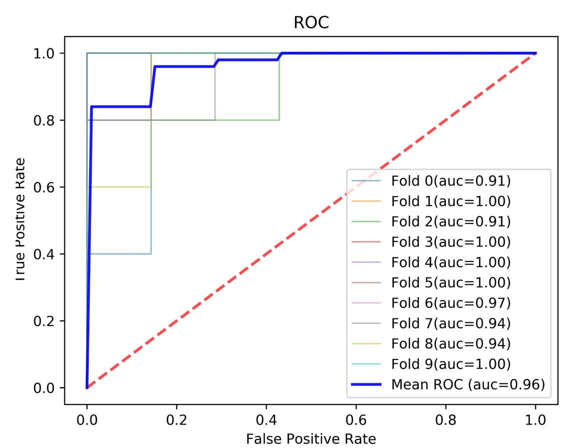


Fig. 5. The ROC curve of SVM classifier, where the blue-black ROC curve is the average of ten ROC curves of 10-folds cross-validation

On the one hand, under the condition of using the same sMRI modality, our proposed method outperformed paper [9]–[11], achieving a accuracy of 89.4%, 8.4% higher than paper [9], 16.3 % higher than paper [11] and 28.3 % higher than paper [10]. It illuminates that the proposed CNN-based method is more effective than other methods just based on traditional machine learning algorithm. CNN plays a significant role in this study considering that it can efficiently extract useful features in different level and reduce the error resulting from incomplete prior hypothesis to some extend. For example, EMCI may have different pathosis from AD, such as degree of lesion, while the prior hypothesis of some researches about EMCI lesion still based on AD. On the contrary, CNN can ignore this difference between EMCI and AD better than prior hypothesis because CNN can automatically extract the most discriminative features no matter what the pathology is. On the other hand, the accuracy of using the features extracted from fine-tuned VGG16 model is much higher than that of using the features from the VGG16 model trained from scratch. The main reason is that VGG16 model trained from scratch can not learn enough features with small dataset, conversely, the VGG16 model transferring pre-trained weights trained by much bigger source-domain dataset has good generalization for small dataset, which can extract many general features of target-domain data from the first few layers without updating weights. It also illustrates that model trained on mass of general images can be fine-tuned by few MRI images to extract specific target-domain features.

In addition, LASSO algorithm is a key step to improve classification accuracy in this study. As shown in Table II, the LASSO algorithm improve the accuracy by 9.4%. It demonstrates that the LASSO algorithm can effectively eliminate redundant features that affect the reasonable distribution of data in SVM feature space. It also explains that there are only subtle but manifest changes for EMCI subjects; therefore these changes are sensitive to vast redundant features.

V. Conclusion

The incidence of Alzheimer's disease is increasing rapidly every year. It is urgent to find effective methods to delay and prevent AD. Diagnosis of EMCI is helpful for early intervention of AD. We reviewed previous researches about the diagnosis of EMCI and proposed a more effective method combining CNN and traditional machine learning approaches to distinguish EMCI from NC. The results suggest that the classification performance is significantly improved comparing with the previous researches, and the average accuracy achieves an unprecedented 89.4%. The excellent classification performance is attributed to the rationality of proposed method. Two factors are identified as being potentially important: 1) Transfer learning based on CNN can greatly enhance the learning ability of small dataset and helps excavate more target-domain high-level features. Additionally, another advantage of transfer learning is the less training time due to the usage of pre-trained model realizing faster convergence with less data. 2) LASSO is a vital step in our study, which can deal with a small number of subjects with high-dimensional features and get rid of a large number of redundant features.

CNN gradually becomes an important tool of analyzing medical image with small dataset. It not only can reduce the need of prior knowledge, but also improve diagnosis accuracy and enhance model's robustness. However, the results of classification using CNN method in this study still have a certain gap with clinical diagnosis. The possible reason is that there are no significative difference between some EMCI subjects and NC in brain structure reflected from sMRI. Therefore, we will make full use of CNN method and other neuroimaging technique to further improve the classification performance for EMCI diagnosis in the following study. In addition, multi-modality fusion is also a noteworthy research direction.

VI. acknowledgment

We are grateful for the continued support from the National Natural Science Foundation of China (No. 81960312) and the Natural Science Foundation of Guangdong Province (No.2018A030310511).

Data collection and sharing for this study was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Instiutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012).

References

[1] A. Association, "2014 Alzheimer's disease facts and figures," *Alzheimers & Dementia the Journal of the Alzheimers Association*, vol. 10, no. 2, pp. e47–e92, 2014.

[2] A. Association, "2018 Alzheimer's disease facts and figures," *Alzheimers & Dementia*, vol. 14, no. 3, pp. 367–429, 2018.

[3] L. S. Schneider, "Mild cognitive impairment," *Lancet*, vol. 367, no. 9518, p. 1262, 2006.

[4] E. Y. Lim, "Comparison of diagnostic accuracy of volumetry or diffusion tensor imaging in mild cognitive impairment," *Alzheimers & Dementia the Journal of the Alzheimers Association*, vol. 12, no. 7, pp. P728–P729, 2016.

[5] W. Wen and W. Li, "The mathematical model of mild cognitive impairment based on brain network analysis," in *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference*, March 2019, pp. 2234–2238.

[6] Y. Zhang, Z. Dong, A. Liu, S. Wang, G. Ji, Z. Zheng, and J. Yang, "Magnetic resonance brain image classification via stationary wavelet transform and generalized eigenvalue proximal support vector machine," *Journal of Medical Imaging and Health Informatics*, vol. 5, no. 7, 2015.

[7] G. Prasad, S. H. Joshi, T. M. Nir, A. W. Toga, and P. M. Thompson, "Brain connectivity and novel network measures for Alzheimer's disease classification," *Neurobiology of Aging*, vol. 36, no. 3, pp. S121–S131, 2015.

[8] E. Kang and H. I. Suk, "Probabilistic source separation on resting-state fMRI and its use for early MCI identification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2018.

[9] R. Raeper, A. Lisowska, and I. Rekik, "Cooperative correlational and discriminative ensemble classifier learning for early dementia diagnosis using morphological brain multiplexes," *IEEE Access*, vol. PP, no. 99, pp. 1–1, 2018.

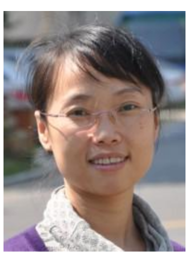
[10] P. Forouzannezhad, A. Abbaspour, C. Li, M. Cabrerizo, and M. Adjouadi, "A deep neural network approach for early diagnosis of mild cognitive impairment using multiple features," in *2018 17th IEEE International Conference on Machine Learning and Applications*, Dec 2018, pp. 1341–1346.

[11] P. Forouzannezhad, A. Abbaspour, M. Cabrerizo, and M. Adjouadi, "Early diagnosis of mild cognitive impairment using random forest feature selection," in *2018 IEEE Biomedical Circuits and Systems Conference*, Oct 2018, pp. 1–4.

[12] T. Kam, H. Zhang, Z. Jiao, and D. Shen, "Deep learning of static and dynamic brain functional networks for early MCI detection," *IEEE Transactions on Medical Imaging*, pp. 1–1, 2019.

[13] Z. Qi, M. Goryawala, M. Cabrerizo, W. Barker, D. Loewenstein, R. Duara, and M. Adjouadi, "Multivariate analysis of structural mri and pet (FDG and 18F-AV-45) for Alzheimer's disease and its prodromal stages," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2014, pp. 1051–1054, 2014.

[14] Y. Cabrera-Len, P. G. Bez, J. Ruiz-Alzola, and C. P. Surez-Araujo, “Classification of mild cognitive impairment stages using machine learning methods,” in *2018 IEEE 22nd International Conference on Intelligent Engineering Systems*, June 2018, pp. 000 067–000 072.
[15] J. Islam and Y. Zhang, “An Ensemble of Deep Convolutional Neural Networks for Alzheimer’s Disease Detection and Classification,” *arXiv e-prints*, p. arXiv:1712.01675, Dec 2017.
[16] L. Yue, X. Gong, J. Li, H. Ji, M. Li, and A. K. Nandi, “Hierarchical feature extraction for early Alzheimers disease diagnosis,” *IEEE Access*, vol. 7, pp. 93 752–93 760, 2019.
[17] L. Yue, X. Gong, K. Chen, M. Mao, J. Li, A. K. Nandi, and M. Li, “Auto-detection of Alzheimer’s disease using deep convolutional neural networks,” in *2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*, July 2018, pp. 228–234.
[18] M. Puranik, H. Shah, K. Shah, and S. Bagul, “Intelligent Alzheimer’s detector using deep learning,” in *2018 Second International Conference on Intelligent Computing and Control Systems*, 06 2018, pp. 318–323.
[19] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *Computer Science*, 2014.
[20] S. P. K. Karri, D. Chakraborty, and J. Chatterjee, “Demnet: A convolutional neural network for the detection of alzheimers disease and mild cognitive impairment,” *Biomed. Opt. Express*, vol. 8, no. 2, pp. 579–592, Feb 2017. [Online]. Available: http://www.osapublishing.org/boe/abstract.cfm?URI=boe-8-2-579
[21] S. J. Pan and Y. Qiang, “A survey on transfer learning,” *IEEE Transactions on Knowledge & Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
[22] M. Hon and N. Khan, “Towards Alzheimer’s disease classification through transfer learning,” in *2017 IEEE International Conference on Bioinformatics and Biomedicine*, 11 2017.
[23] Z. Jiao, Z. Xia, X. Ming, C. Cheng, and S. Wang, “Multi-scale feature combination of brain functional network for eMCI classification,” *IEEE Access*, vol. PP, pp. 1–1, 06 2019.
[24] J. A. K. Suykens, “Support Vector Machines: A nonlinear modelling and control perspective,” *European Journal of Control*, vol. 7, no. 2-3, pp. 311–327, 2001.
[25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.



Li Kang (M’87) received the Ph.D degree in signal and Information Processing from Xidian University in 2008. She was a Postdoc with the Optical Engineering at Shenzhen University. She is now a lecturer in College of Information Engineering, Shenzhen University, China. Her research interests include medical image processing and intelligent information processing.

Jingwan Jiang received the B.S. degree in Electronic Information Science and Technology from Wuhan University of Technology in 2018. She is now a postgraduate in College of Information Engineering, Shenzhen University, China. Her research interests include medical image processing and intelligent information processing.

Jianjun Huang (M’02) received the Ph.D degree in Signal and Information Processing from Xidian University in 1997. He was a Postdoc in Computer Sciences with Northwestern Polytechnical University from 1997 to 1999. He is now a professor with College of Information Engineering, Shenzhen University, China. His current research interests include intelligent information processing, computer vision and medical image processing.

Tijiang Zhang received the Ph.D.degree of Medical Imaging in West China Medical School of Sichuan University in 2010. The director of Medical Imaging Center of Guizhou Province and Department of Radiology, the Affiliated Hospital of Zunyi Medical University. The 11th batch Youth technology talented person of Guizhou province. He is now professor with Radiology, the Affiliated Hospital of Zunyi Medical University. His current research interests include imaging diagnosis of Central nervous system and clinical application of high field MRI.