Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2019.Doi Number

Hierarchical Feature Extraction for Early Alzheimer's Disease Diagnosis

Lulu Yue [1], Xiaoliang Gong [1,*], Jie Li[1], Hongfei Ji[1], Maozhen Li[2] and Asoke K. Nandi[2]

[1]Department of Computer Science and Technology, Tongji University, Shanghai, 201804, P.R.China
[2]Department of Electronic and Computer Engineering, Brunel University London, Uxbridge UB8 3PH, UK

*Corresponding author: Xiaoliang Gong (e-mail: gxllshsh@163.com).

This work was supported by the Science and Technology Commission of Shanghai Municipality under grants (16JC1401300, 7ZR1431600, 18ZR1442700), the Shanghai Sailing Program (16YF1415300), Special Fund for Basic Scientific Research Business Expenses of Central Colleges and Universities (22120180542), and the Fundamental Research Funds for the Central Universities.

ABSTRACT Mild cognitive impairment (MCI) is the early stage of Alzheimer's disease (AD). In this article, we propose a novel voxel-based hierarchical feature extraction (VHFE) method for the early AD diagnosis. First, we parcellate the whole brain into 90 regions of interests (ROIs) based on an Automated Anatomical Labeling (AAL) template. To split the uninformative data, we select the informative voxels in each ROI with a baseline of their values and arrange them into a vector. Then, the first stage features are selected based on the correlation of the voxels between different groups. Next, the brain feature maps of each subjects made up of the fetched voxels is fed into a convolutional neural network (CNN) to learn the deep hidden features. Finally, to validate the effectiveness of the proposed method, we test it with the subset of the Alzheimer's Disease Neuroimaging (ADNI) database. The testing results demonstrate that the proposed method is robust with promising performance in comparison with the state-of-the-art methods.

Key words: Alzheimer's disease; Convolutional neural network; Hierarchical feature extraction; Mild cognitive impairment.

I. INTRODUCTION

Alzheimer's disease (AD) is one of the most common degenerative brain diseases. There are more than 50 million people in the world, who are suffering from Alzheimer's disease and other dementias [1]. The typical symptoms of AD are a continuous decline in thinking, behavioral and social skills that disrupt a person's ability to function independently [2]. It is both a mental and financial burden on a family if there is an Alzheimer' disease sufferer [3] [4]. With the progress of science and technology, medical health care has helped to increase the average life of the human beings. But in the past 20 years, only two types of drugs were discovered to treat some symptoms of the disease [1]. Mild cognitive impairment (MCI) is a decline in memory or other thinking skills. People who have MCI would face a significant risk of developing dementia. The primary MCI deficit is memory and this condition is more likely to progress to dementia due to Alzheimer's disease. In its early stages, memory loss is mild, but with late-stage Alzheimer's, individuals lose the ability to carry on a conversation and respond to their environment. As a result, it will represent a significant contribution to be able to diagnose Alzheimer's disease at an early stage to help delay deterioration [5].

As a safe, rapid accurate clinical diagnosis method without any harm to human body, Magnetic resonance imaging (MRI) is widely used in clinical diagnosis. In recent years, artificial intelligence has shown great advantages in computer aided diagnosis. We can extract meaningful features from large dimensional MRI images by machine learning methods. Generally, the feature learning methods can be divided into three categories which are regions of interests (ROIs)-based methods, voxels-based methods and patch-based methods [6]. ROIs-based methods extract features in regions that are parcellated based on anatomical or functional atlas. Due to its small data size, it has been widely used in the early research studies [7] [8] [9]. However, in ROIs-based methods, features were extracted based on the overall changes of each ROI where the subtle variations are barely covered. The voxels-based methods can solve this problem because it can figure out the subtle changes in brain. However, voxels-based methods incur a data set of high dimension which is computationally expensive. Patch-based methods have been proposed to make

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2926288, IEEE Access

up for these short comings. Liu et al. proposed a local patch-based subspace ensemble method [10]. The whole brain was segmented into a set of patches. Classifiers were used to learn the optimal sparse representation by randomly select some subsets. And then a feature vector were constructed with the voxel densities. Zhang et al. [11] proposed a landmark-based feature extraction method. The work was divided into two stages. In the first stage, the landmarks were figured out by comparing the local morphological differences. In the second stage, a regression forest was used to find the landmarks in the testing data. The limitation of this method is the number of training data and the error of detecting landmarks may also affect the results [12]. Liu et al. also proposed a landmark-based framework in his article. What different is that he used a multi-instance convolutional neural network (CNN) to learn the representation of each patch. And then the features were concatenated together and fed into another deep 3D CNN model. However, classification results of this method are mostly limited by the number of the training data. Liu et al. proposed a 3-D texture feature learning framework. To learn the best nodes and edge features, multiple kernel classifiers were used. But they only used F-score for feature selection [12]. There are also many scientists using multiple modality data in their researches. Suk et al. proposed to use a multi-modal Deep Boltzmann Machine (DBM) to extract the latent features of the 3-D patches learned from the MRI images and Positron Emission Tomography (PET) data [6]. It worth mention that they made a fusion of the 3-D patches of MRI and PET images so than they can fetch the representations that contains the correlations between the multimode data. Liu et al. proposed to use a stacked auto-encoders to learn the optimal representations of MRI and PET data by randomly hiding one modality in the training set. So that the features can reflect the interactions of the two model of data.

In this article, we propose a novel voxel-based hierarchical feature extraction (VHFE) method. First, we extract the first-level features by calculating the correlation between subjects at a voxel level. Then, the features are processed in the form of feature vectors and fed into a classifier to verify the effectiveness of the features. Next, the morphological variation related features are organized into a brain feature map. To capture the deep hidden features of the whole brain, the brain feature maps are fed into a convolutional neural network to learn the deep global features.

The major contributions of the paper are as follows:

(1) A novel voxel-based hierarchical feature extraction method, which provides to be a more convenient and effective method in AD diagnosis, is proposed.

(2) Feature vectors are made up with voxels that are selected in strict flow and non-registration is needed. Furthermore, the effect of registration error on classification results is avoided.

(3) The proposed method not only greatly reduces the data dimension and calculation cost, but also covers the subtle pathological changes at the voxel level.

The rest of the paper is organized as follows. Section II details the data use in this research and its preprocessing. Section III introduces the proposed method. Section IV evaluates the performance of VHFE and discusses the results. Section V concludes the paper.

II. DATASETS AND PREPROCESSING

We chose two datasets from the ADNI database to confirm the framework proposed in this research. ADNI is a longitudinal multicenter study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of AD (http://adni.loni.usc.edu/).

A. DATASET

All the subjects in this research are selected from the ADNI database. We choose two datasets (ADNI-1 and ADNI-2) here to verify the method proposed in this article.

1) ADNI-1

The ADNI-1 database is composed of three different stages of subjects: normal controllers (NC), mild cognitive impairment (MCI), and AD. Particularly, we chose the structural MRI data which were scanned with 1.5 Tesla SIEMENS nuclear magnetic resonance scanner. Flip Angle is 8.0 degree; Slice thickness of each image is 1.2mm, Echo time (TE) is 3.6ms, inversion time (TI) is 1000.0 ms and repetition time (TR) is 3000.0 ms. All the images were preprocessed by GradWarp and B1 Correction with pro_ADNI_script [14], then processed by ADNI pipeline with nonparametric non-uniform intensity normalization (N3) algorithm for a correction of intensity inhomogeneity [10][14]. Despite the ill-formatted data, there are 1662 volumes remained including 785 NC, 542 MCI, 335 AD. The subject info is detailed in Table1.

2) ADNI-2

The T1 weighted structural images in ADNI-2 were scanned with 3.0 Tesla SIEMENS nuclear magnetic resonance scanner. The image Slice thickness is 1.2 mm, TE is 2.95 ms, TI is 900.0 ms, and TR is 2300.0 ms. The data were preprocessed a little different from that in ADNI-1. First, the images were processed to correct gradient non-linearity distortions [16]. Then, N3 algorithm was also implemented here. Different from the ADNI-1database, there are four categories in ADNI-2 dataset including 1106 NC, 1320 early mild cognitive impairment (EMCI), 987 late mild cognitive impairment (LMCI), and 305 AD. The subject info is detailed in Table2.

TABLE I
DEMOGRAPHIC AND CLINICAL INFORMATION OF ADNI-1

	Number	Age	Gender(Female/Male)	MMSE
NC	785	74.63±3.69	416/369	29.07±1.32
MCI	542	78.86±5.35	193/349	26.56±2.63
AD	335	78.56±5.34	156/180	23.84±2.10

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2926288, IEEE Access

TABLE 2
DEMOGRAPHIC AND CLINICAL INFORMATION OF ADNI -2

	Number	Age	Gender(Female/Male)	MMSE
NC	1106	74.63±3.69	554/552	29.10±1.25
EMCI	1583	76.86±4.97	570/1013	28.37±1.48
LMCI	1304	76.53±5.35	639/665	27.19±2.23
AD	366	78.58±5.38	138/228	21.84±4.10

B. PREPROCESSING

As mentioned above, in order to verify the validity of the method, we selected subjects form two subsets from the ADNI database. Then a strictly preprocessing pipeline was implemented. Firstly, the T1 images were normalized to a template space and segmented into gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF). After the quality check step, we smoothed the GM images with the smooth module in SPM12. We preprocessed all the data with voxel-based morphometry (VBM8) [17] which is a neuroimaging analysis technique that uses statistical methods of statistical parameter mapping to study local differences in brain anatomy [18]. Then, we used AAL [19] to segment the volume into 126 regions of interests (ROIs). After throwing away the regions belong to the cerebellar, we got 90 regions for every subject [20].

III. PROPOSED METHOD

In this section, we proposed a VHFE method to mine inner region abnormalities in structural MRI images. The data processing flow chart is demonstrated in Figure1. Firstly, we preprocessed all the structural MRI images as described above. Then we picked all the voxels in each region and fed them into a matrix respectively. The ROIs were parcellated based on the AAL template and it results in there being different number of voxels in each region. We used the Kendall's correlation coefficient to select the most irrelevant voxels between different groups of subjects as the feature of the first stage. Fifty voxels were selected from each region. Then all the voxels of each region make up the whole brain map. The brain map were then fed into CNN to learn the deep hidden feature inner or between subjects as the feature of the second stage. Finally, the result of a softmax classifier is used to evaluate the efficiency of the proposed framework. The schematic diagram is shown in Figure1.

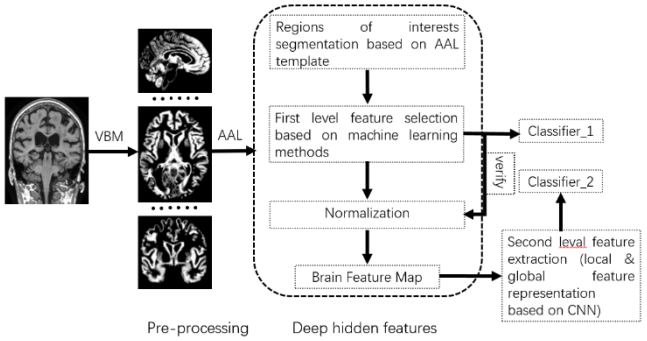

FIGURE 1. The proposed data processing flow chart.

1) INNER-REGION FEATURE SELECTION

After the preprocessing procedure, the data remained in the GM volume stands for the voxel intensity. Due to the AAL template we used to parcellate the GM into 90 ROIs. The number of voxels differ in each of these ROIs. Some contains only a few hundreds of voxels while some can be more than ten thousand. The original methods used to average the data in each ROIs and then fed them into an SVM classifier to make judgements. But here, we resliced all the voxels in each ROIs into a vector with the same rule, according to the scanning order and the row each voxel was in. That is, if there is γ_n voxels in the n-th ROIs, $\gamma_n \in \{\gamma_1, \gamma_2, \gamma_3, \dots, \gamma_n\}_{n\in N}$. N stands for the number of the voxels in each ROI. As the feature extracted in the first stage are used to make up the whole brain feature map, we chose the number of ten percent of the voxels in the smallest ROI as the baseline for the number of features extracted from each ROI. Finally, we used Pearson correlation, Kendall's rank correlation and Spearman correlation to figure out 50 of the most irrelevant voxels in the ROIs to figure out the most irrelevant voxels in each ROI among groups. For the n-th voxel, we also construct a feature matrix $\{\Gamma_1, \Gamma_2, \Gamma_3, \dots, \Gamma_i\}_{i\in I}$, where i represents the number of subjects in each groups. We used Kendall's rank correlation to pick out 50 of the most irrelevant voxels in the ROIs. In statistics, the Kendall rank correlation coefficient, commonly referred to as Kendall's tau (τ) coefficient, is a statistic used to measure the association between two measured quantities. Comparing with the Pearson correlation coefficient which can only measures linear dependence relations, the Kendall's correlation coefficient, is more suited for use in image processing where stationarity cannot usually be advocated. The Pearson correlation and Spearman correlation were also used to validate the assumption.

We used a random forest (RF) regression framework to check the features we captured from each ROI in the first stage. First, the average values of each of the ROIs were put together for a new feature vector $\{\delta_1, \delta_2, \dots, \delta_k, \dots, \delta_{90}\}_{k\in 1,2,\dots,90}$. Then, the new vector was labeled and fed into a random forest (RF) regression framework to check out the effectiveness of the selected voxels. The result can be seen from Table 3 and Table 4. The fusion of the top 50 most irrelevant voxels in the ROIs made up the whole brain map for each subjects. Then the brain map were labeled and then fed into the convolutional neural network to learn the deep hidden features of the subjects.

2) BRAIN MAP FORMULATION AND CLASSIFICATION

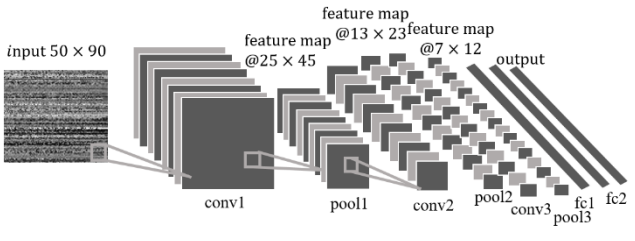

FIGURE 2. The convolutional neural network.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2926288, IEEE Access

CNN is a kind of deep, feed-forward neural network. In the past years, CNN has shown its superiority in feature learning especially for large dimensional data. As the traditional neural network, CNN is composed of the input layer, the output layer, convolutional layers and subsampling layers. Each layer contains different number of nodes with learnable weights and bias. Each neuron performs a dot between inputs and weights. The results of the operation is determined by different types of activation functions. The pooling layer here averaged the sampled data for dimensionality reduction. CNNs exploit spatially-local correlation by enforcing a local connectivity pattern between neurons of adjacent layers. Weight sharing greatly reduces the number of weights used for training. In each convolution layer, the outputs of its previous layer are convolved with a learnable kernels. Then the feature map was formed by the activation function as the outputs. Generally, the formula can be described as

$$y_\tau^\iota = f\left(\sum_{\ell \in N_\tau} y_\ell^{\iota-1} * w_{\ell\tau}^\iota + b_\tau^\iota\right), \qquad (1)$$

where N_j represents the number of the input maps and f is the activation function.

The pooling layer reduces the dimensionality of the inputs by a down-sampling operation. The subsampling layer is to divide the feature map of the output of the convolutional layer into several regions, each region is represented by the value of the region. More formally,

$$y_j^\ell = f\left(\beta_i^\ell down\left(y_j^{\ell-1}\right) + b_j^\ell\right), \qquad (2)$$

where f is an activation function and $down(\cdot)$ represents the function of the sub-sampling.

The backpropagation technique here uses a feedforward structure to propagate errors in the neural network in order to adapt the weights. Backpropagation is a method of achieving gradient descent in neural networks. The output layer error is defined as

$$\delta_j^{(\ell)} = a_j^{(\ell)} - y_j, \qquad (3)$$

where hidden layer error signal is written as

$$\delta^{(i)} = (\theta^{(i)})^T \delta^{(i+1)} * \Delta a^{(i)} \qquad (4)$$

where $\theta^{(i)}$ represents weights of layer i. The $\delta^{(i)}$ represents the back-propagated error signal, which is used to update the activation values in layer i and $\Delta a^{(i)}$ represents the gradients of the activation function in layer i.

The CNN we implemented in this article is shown in Figure 2 which included three convolutional and three sub-sampling layers. The Linear Unit (Relu) activation function was adopted in each convolutional layer. After each pooling layer, we set fully connected layers behind the last pooling layer. A 64-bit 16GB RAM PC with a 8GB GTX1080 GPU was used in our test. We set the learning rate to 0.5 and the threshold we set for the loss function is 0.001.

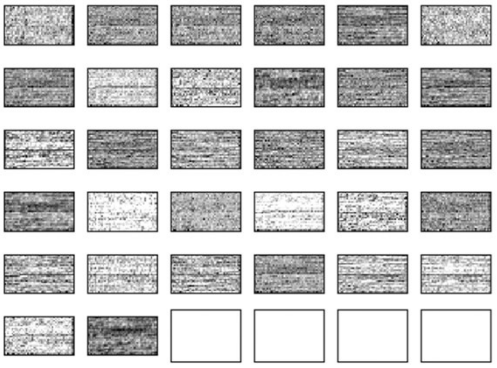

FIGURE 3. The features fetched by the first convolutional layer on classification of AD/NC in ADNI-1.

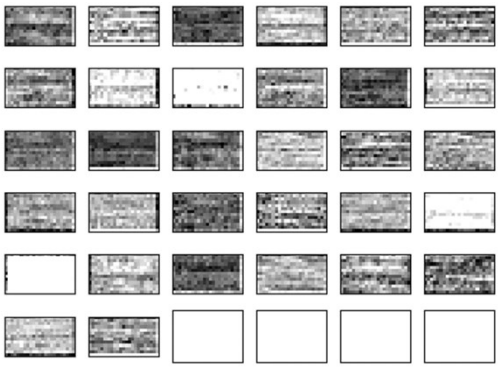

FIGURE 4. The feature fetched by the second convolutional layer on classification of AD/NC in ADNI-1.

FIGURE 5. The features fetched by the third convolutional layer on classification of AD/NC in ADNI-1.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2926288, IEEE Access

IV. RESULTS AND DISCUSSION

In order to validate the proposed method in this article, we download data from the ADNI-1 and ADNI-2 dataset respectively. In ADNI-1 dataset, there are three categories of subject. So we separate them into three groups to do the binary classification: AD vs NC, AD vs MCI, MCI vs NC. In the ADNI-2 dataset, there are four categories of subjects. And then there should be six matched groups: NC vs EMCI, NC vs LMCI, NC vs AD, EMCI vs LMCI, EMCI vs AD, LMCI vs AD. In the experiment, the each dataset was randomly shuffled and then partitioned into two part. We randomly selected 20 percent of each groups as testing data which were absolutely separated from the training data. In order to insure the robustness of the result, the cross-validation was applied. Each time, the rest of data was divided into 5-folds. Among them, one fold was taken as the validation data to make sure that the experiment is not locally optimal and the other data used for training. The final result is the average of ten repeated tests.

A. CLASSIFICATION RESULTS

Feature maps of each convolutional layer are shown in Figure 3, 4, 5. Table 3 and Table 4 shows the results of three different inputs based on Pearson correlation, Kendall correlation, Spearman correlation respectively and the results of the baseline on different groups. The column named "RF+mean" refers to the results of the baseline. From the Table 3 and Table 4 we can see that the most irrelevant voxels we selected based on three correlation coefficients provide a better result than the baseline. Specifically, the Kendall's rank correlation increase 8% on classifying AD and MCI, more than 20% in classifying AD from NC, and almost 16% in classifying NC from MCI compared to the baseline. The Pearson correlation and Spearman correlation also performed a much higher classification result on ADNI-1. On ADNI-2, the Kendall's rank correlation also increase the accuracy much more than other feature selection methods. Specially, it increases 10.5% (AD vs NC), 8% (AD vs EMCI), 6.5% (NC vs EMCI), 15.9% (NC vs LMCI) and 7% (EMCI vs LMCI) compared to the baseline. Even though Pearson correlation and Spearman correlation offer better performance than the original method, the Kendall's rank correlation seems better in most instances. The receiver operating characteristic (ROC) curves for the classification of the features extracted by the Kendall's rank correlation methods in different groups were shown in Fig 6. The true positive rate (TPR) stands for the proportion of positive instances identified by the classifier to all positive instances. The false positive rate (FPR) stands for the proportion of all negative instances where the classifier mistakenly considers a positive class. The area under the curve (AUC) is 0.97 in classifying AD from NC, and we also got 0.9 and 0.8 when identifying MCI from NC and AD respectively, which proves that the feature we extracted is positive.

TABLE 3
PERFORMANCE COMPARISON ON THREE DIFFERENT CLASSIFICATION TASKS WITH DIFFERENT FEATURE SELECTION METHODS IN ADNI-1

ADNI1	RF + mean	RF +Pearson	RF + Kendall	RF +Spearman	RF +Pearson + Kendall + Spearman
AD vs MCI	62.9±3.1	61.0±11.3	70.9±7.5	69.5±8.0	70.9±9.8
AD vs NC	69.4±1.8	87.2±7.9	90.9±6.8	87.0±9.8	89.7±6.1
NC vs MCI	59.4±1.9	77.2±11.8	76.5±11.8	81.1±8.1	75.3±8.9

TABLE 4
PERFORMANCE COMPARISON ON THREE DIFFERENT CLASSIFICATION TASKS WITH DIFFERENT FEATURE SELECTION METHODS IN ADNI-2

ADNI2	RF + mean	RF + Pearson	RF + Kendall	RF + Spearman	RF + Pearson + Kendall + Spearman
AD vs NC	74.9±1.6	83.9±7.2	85.4±7.2	83.4±7.9	78.5±15.3
AD vs LMCI	76.2±1.9	68.7±7.1	66.5±10.3	66.8±9.5	71.3±7.7
AD vs EMCI	80.7±1.8	83.6±8.8	88.8±4.4	83.4±9.8	81.8±13.7
NC vs EMCI	59.6±1.2	60.8±6.0	66.1±6.0	62.4±7.5	59.4±8.7
NC vs LMCI	52.1±1.4	66.5±7.4	68.0±7.3	67.1±9.2	66.6±9.6
EMCI vs LMCI	57.3±1.6	64.6±7.4	64.3±8.1	60.5±9.5	62.5±9.0

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2926288, IEEE Access

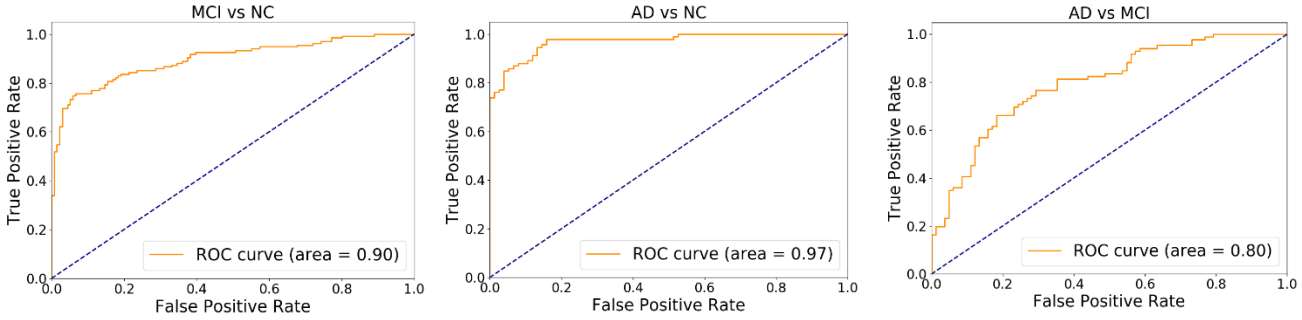

FIGURE 6. **Receiver Operating Characteristic (ROC) curves for RF + Kendall in classifying AD from NC on ADNI-1.**

The features we selected at the first stage were validated to be effective. So we fused all the regions together to construct the brain feature map. Then we used a convolutional neural network to learn the voxel-based deep hidden features inner and between each group. The results can be seen from Table 5 and Table 6. In Table 5, the column named "CNN + Raw" means that the input data was just preprocessed as described in session III. Then we resliced the three-dimensional GM images into a series of two-dimensional images. Then these images were fed into the convolutional neuro network to learn deep hidden features as well. Specially, the number of subjects remained constant in all of these competing methods but, due to the different feature selection method, the number of images in the method lists in the column" CNN + Raw" is much more than the others. As a result, the computation time of the proposed method is almost 57 seconds. However, it takes almost 20 minutes when put the resliced GM images in the CNN framework.

As shown in Table 5, the proposed method obtains a result of 97.8% (AD vs MCI), 99.7% (AD vs NC), and 97.7% (NC vs MCI) with the Kendall's rank correlation was done in the first phase. We can see that when the data were selected using the Spearman correlation at the first phase, we even got a 100% accuracy when classifying AD from NC. The confusion matrix of each group which processed by the Kendall's rank correlation algorithm can be seen from Fig7. The first column in the first row and the second column in the second row stands for the number of Represent the number of subjects which were correctly classified. It means the accuracy is higher when it is getting yellow. Table 6 shows that the proposed method shows a stable advantage on ADNI-2. It enhanced the accuracies by 1.7% (AD vs NC), 2.66% (AD vs LMCI), 1.99% (AD vs EMCI), 4.16% (NC vs EMCI), 3.97% (NC vs LMCI) and 3.16% (EMCI vs LMCI) compared with the method we proposed in the previous article [18].

TABLE 5
PERFORMANCE COMPARISON OF THE PROPOSED METHOD WITH DIFFERENT FEATURE SELECTION METHODS IN ADNI-1

ADNI1	CNN + Raw (%)	CNN + Pearson (%)	CNN + Kendall (%)	CNN + Spearman (%)
AD vs MCI	93.89±4.40	96.00±2.90	97.80±1.30	98.60±0.02
AD vs NC	95.44±0.40	99.50±0.80	99.70±0.70	100.00±0.00
NC vs MCI	95.38±0.30	98.80±1.20	98.90±1.00	96.90±0.80

TABLE 6
PERFORMANCE COMPARISON OF THE PROPOSED METHOD WITH DIFFERENT FEATURE SELECTION METHODS ON ADNI-2

ADNI2	CNN + Raw (%)	CNN + Pearson (%)	CNN + Kendall (%)	CNN + Spearman (%)
AD vs NC	96.91±0.01	99.40±1.10	98.60±2.90	98.30±0.50
AD vs LMCI	97.14±0.01	97.40±0.70	99.80±0.50	98.60±0.50
AD vs EMCI	97.81±0.00	100.00±0.00	99.80±0.50	99.50±0.80
NC vs EMCI	95.44±0.08	99.00±0. 80	99.60±0.40	99.10±0.50
NC vs LMCI	94.43±0.17	97.80±0. 50	98.40±0.30	99.30±0.60
EMCI vs LMCI	94.84±0.03	96.70±0. 60	98.00±0.70	97.70±0.50

TABLE 7
CLASSIFICATION PERFORMANCE FOR DIFFERENT GROUPS ON ADNI-1

ADNI1	Accuracy score (%)	Precision score (%)	Recall score (%)	F1 score (%)
AD vs MCI	97.2±2.1	96.1±2.8	98.4±2.2	97.2±2.0
AD vs NC	99.4±1.5	98.8±2.8	100.0±0.0	99.4±1.4
NC vs MCI	98.9±1.0	99.4±0.9	98.5±1.3	98.9±1.0

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2926288, IEEE Access

TABLE 8
CLASSIFICATION PERFORMANCE FOR DIFFERENT GROUPS ON ADNI-2

ADNI2	Accuracy score (%)	Precision score (%)	Recall score (%)	F1 score (%)
AD vs NC	98.6±0.5	100±0.0	97.2±1.0	98.6±0.5
AD vs LMCI	99.7±0.7	99.7±1.0	99.7±1.0	99.7±0.7
AD vs EMCI	100±0.0	100.0±0.0	100.0±0.0	100.0±0.0
NC vs EMCI	99.7±0.3	99.9±0.2	99.5±0.5	99.7±0.3
NC vs LMCI	98.5±0.5	99.0±0.4	98.0±0.9	98.5±0.5
EMCI vs LMCI	98.0±0.6	98.9±0.01	97.2±0.8	98.0±0.6

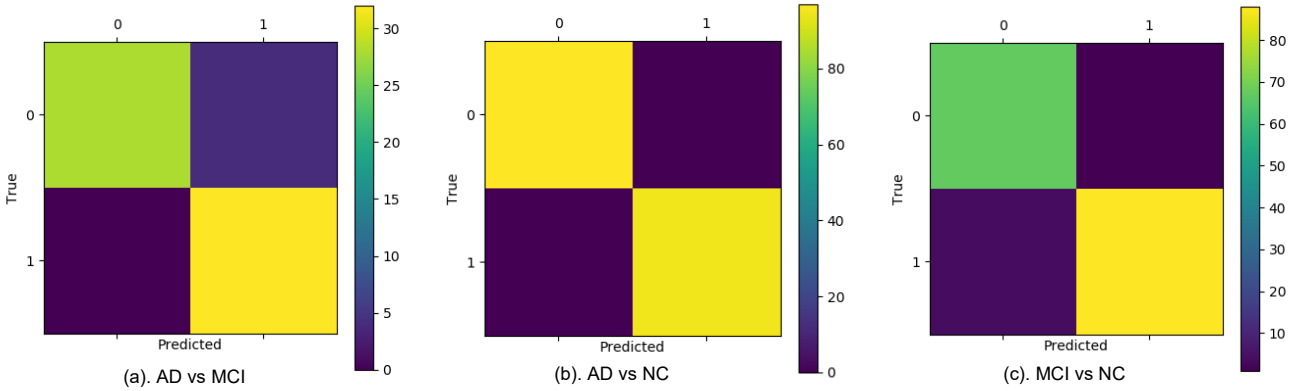

FIGURE 7. The confusion matrixes with three binary classifications on ADNI-1.

Table 7 and Table 8 record the performance of the proposed method on ADNI-1 and ADNI-2 respectively with the pre-feature selection method of Kendall correlation. Each experiment was repeated ten times and the results here are the average value of ten tests. It is worth noting that, the proposed method showed an outstanding performance both in distinguishing MCI from NC and EMCI from NC. Table 8 shows that our proposed method performed best on the three kind of binary classification on ADNI-1. Specially, we got an accuracy improvement of 12.55% compared to the state-of-the-art methods in classifying MCI from NC. It is very important and meaningful for diagnosing MCI from NC at an early stage. Also, we got 4.5% and 6.95% improvement in classifying NC vs MCI and AD vs MCI respectively.

TABLE 8. PERFORMANCE COMPARISON OF THE PROPOSED METHOD WITH THE STATE-OF-THE-ART METHODS ON ADNI-1

ADNI1	AD vs MCI	AD vs NC	NC vs MCI
Chupin et al. [22]	73.48	80.51	71.94
Ahmed et al. [23]	74.51	86.40	76.29
Suk et al. [24]	88.98	93.05	83.67
Khedher et al. [25]	84.59	88.96	82.41
Dai et al. [26]	85.92	90.81	81.92
Liu et al. [27]	90.85	95.24	86.35
Proposed method	97.80±1.30	99.70±0.70	98.90±1.00

B. DISCUSSION AND LIMITATIONS

Rigorous comparison and verification were done to verify the effectiveness of our proposed method: (1) At the first step, after the preprocessed data was segmented into GM, WM and CSF, the GM images were parcellated into 90 regions of interests (ROIs). We take the average of each ROIs as the baseline, which means that the data named "ROI-mean" stands for the one no feature selection was done. Then we picked out 50 most irrelevant voxels in each ROI, and take the average data of them to feed into our trained random forest model to judge the validation of the features. It should be emphasized that we parcellated the GM images based on the AAL template. That leads to the number of voxels in each ROI differs one from the other, so we take the average value of all the voxels as the baseline. To be contrast, we averaged the selected 50 most irrelevant voxels as well. Table 2 and Table 3 detailed the advantages of the selected features. (2) To catch the most typical features, we calculated three different correlation coefficients between each group. The fused feature was extracted out at the same time. (3) Our ultimate objectives were to construct the whole brain map and extract the hierarchical features within and between the subjects. Table 4 and Table 5 show the result of the proposed method with three different kinds of correlation coefficients. (4) Finally, we compared the proposed method with six state-of-the-art methods.

Compared to the traditional ROIs-based methods [28], the proposed VHFE method can capture more subtle changes in each ROI. Not the same as the conventional voxels-based methods, a dimensionality reduction was done after a data driven distinguish feature learning [29] [30]. The first-level feature we extracted not only contains the voxel-level subtle differences between subjects, but also maintained the anatomically functional integrity with the ROIs-level dimensions [31] [32]. Besides, unlike the patch-level methods proposed by Suk and Shen etc., there is no need for registration in our VHFE method. Therefore, errors caused by registration of the test data based on the location of landmarks

are avoided [11] [12]. The hierarchical feature extraction method we proposed can not only capture the local features in each ROI by the feature extraction method in the first stage. In the second stage, the brain feature map can also help learn the global distinct information among different groups. However, there is still much to be improved. First, the features we selected in the first stage only compared the relationship between groups, we can also take the inner-relationships in ROIs into consideration. Secondly, we did not take the complementarity between multimodal data into consideration and our future work should be try to fix on this point. Thirdly, we will try to test and refine our approach on multiple types of data to improve the universality of the approach.

V. CONCLUSION

In this article, we proposed a VHFE method by two stage of procedure. In the first stage we selected the most irrelevant voxels in each ROIs to construct a feature vector. Then the feature vectors made up the brain feature map used for learning deep hidden features inner and between subjects. Specifically, we proposed to find the most informative voxels as the presentation of each ROI. The error caused by matching the position of voxel in the test phase is avoided. In the second stage the CNN can help figure out the subtle changes in deep hidden levels. We selected two subsets of ADNI database to verify our proposed method. The results of the proposed method showed significantly better performance than those from the state-of-the-art methods.

REFERENCES

[1] Alzheimer's Disease International, "World Alzheimer Report 2018 The state of the art of dementia research: New frontiers," Published by Alzheimer's Disease International (ADI), London. September 2018.

[2] R. J. Perrin, A. M. Fagan, D. M. Holtzman, "Multimodal techniques for diagnosis and prognosis of Alzheimer's disease," NATURE, Vol 461|15, pp.916-922, October 2009

[3] J. Barnesa, B. C. Dickerson, C. Frost, L. C. Jiskoot, D. Wolk, W. M. G. Flier, "Alzheimer's disease first symptoms are age dependent: Evidence from the NACC dataset," Volume 11, Issue 11, November 2015, Pages 1349-1357.

[4] P. Maresova, H. Mohelska, J. Dolejs, and K. Kuca, "Socio-economic aspects of Alzheimer's disease," Curr. Alzheimer Res., vol. 12, no. 9, pp. 903–911, 2015.

[5] R. A. Sperling, P. S. Aisen, L. A. Beckett, D. A. Bennett, S. Craft, A. M. Fagan, T. Iwatsubo, C. R. Jack Jr., J. Kaye, T. J. Montine, D. C. Park, E. M. Reiman, C. C. Rowe, E. Siemers, Y. Stern, K. Yaffe, M, C. Carrillo, B. Thies,…C. H. Phelps, "Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease," Alzheimer's Dementia, vol. 7, no. 3, pp. 280–292,May 2011

[6] H. I. Suk, S. Lee, D. Shen "Hierarchical feature representation and multimodal fusion with deeplearning for AD/MCI diagnosis," NeuroImage 101,569–582, 2014.

[7] H. I. Suk, D. Shen, "Deep learning-based feature representation for AD/MCI classification, " Medical Image Computing and Computer-Assisted Intervention, Lecture Notes in Computer Science, vol. 8150, pp. 583–590. MICCAI,2013.

[8] C. Davatzikos, P. Bhatt, L. M. Shaw, K. N. Batmanghelich,, J. Q. Trojanowski, "Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification," Neurobiol Aging 32(12):2322.e19–2322.e27, 2011

[9] H. I. Suk, S. W. Lee, "A novel bayesian framework for discriminative feature extraction in brain-computer interfaces." IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(2), 286-299. 2013.

[10] M. Liu, D. Zhang, D. Shen, "Ensemble sparse classification of Alzheimer's disease," NeuroImage, vol 60, 1106–1116, 2012.

[11] J. Zhang, Y. Gao, Y. Gao, D. Shen, "Detecting Anatomical Landmarks for Fast Alzheimer's Disease Diagnosis," IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 35, NO. 12, DECEMBER 2016

[12] M. Liu, J. Zhang, E. Adeli, D. Shen, "Landmark-based deep multi-instance learning for brain disease diagnosis," Medical Image Analysis 43 (2018) 157–168.

[13] J. Liu, J. Wang, B. Hu, F. Wu, and Y. Pan "Alzheimer's Disease Classification Based on Individual Hierarchical Networks Constructed With 3-D Texture Features," IEEE Transactions on Nanobioscience, VOL. 16, NO. 6, SEPTEMBER 2017.

[14] S. Liu, S. Liu, W. Cai, H. Che, S. Pujol, R. Kikinis, D. Feng, M. J. Fulham "Multimodal Neuroimaging Feature Learning for Multiclass Diagnosis of Alzheimer's Disease." IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 62, NO. 4, APRIL 2015.

[15] J. G. Sled, A. P. Zijdenbos, A. C. Evans. "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," IEEE Trans Med Imaging. 1998; 17(1):87–97. [PubMed: 9617910]

[16] M. J. Hardonk, F. W. Dijkhuis, T. J. Haarsma, J. Koudstaal, W. A. Huijbers,"Application of enzymehistochemical methods to isolated subcellular fractions and to sucrose-ficoll density gradients. A contribution to the comparison of histochemical and biochemical data." Histochemistry. 53(2):165-81. Aug 1, 1977;

[17] J. Ashburner and K. J. Friston, "Voxel-Based Morphometry—The Methods", NeuroImage11, 805–821 (2000), doi:10.1006/nimg. 2000.0582.

[18] L. Yue, X. Gong, K. Chen, M. Mao, J. Li, A. K. Nandi , M. Li" Auto-Detection of Alzheimer's Disease Using Deep Convolutional Neural Networks," 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD),228-234.

[19] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, M. Joliot, "Automated, Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain", NeuroImage, 15, 273–289 (2002).

[20] H. I. Suk, S. W. Lee, D. Shen, "Latent feature representation with stacked auto-encoder for AD/MCI diagnosis," Brain Struct Funct (2015) 220:841–859.

[21] X. Zhu, H. I. Suk, L. Wang, S. W. Lee, D. Shen, "A novel relational regularization feature selection method for joint regression and classification in AD diagnosis," Medical Image Analysis 38 (2017) 205–214.

[22] M. Chupin, E. G´erardin, R. Cuingnet, C. Boutet, L. Lemieux,S. Leh´ericy, H. Benali, L. Garnero, and O. Colliot, "Fully automatic hippocampus segmentation and classification in alzheimer's disease and mild cognitive impairment applied on data from adni,"Hippocampus, vol. 19, no. 6, pp. 579–587, 2009.

[23] O. B. Ahmed, J. Benois-Pineau, M. Allard, C. B. Amar, G. Catheline, A. D. N. Initiative et al., "Classification of alzheimers disease subjects from mri using hippocampal visual features," Multimedia Tools and Applications, vol. 74, no. 4, pp. 1249–1266, 2015.

[24] H. I. Suk, S. W. Lee, D. Shen, A. D. N. Initiative, "Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis," NeuroImage, vol. 101, pp. 569–582, 2014.

[25] D. Dai, H. He, J. T. Vogelstein, and Z. Hou, "Accurate prediction of ad patients using cortical thickness networks," Machine vision and applications, vol. 24, no. 7, pp. 1445–1457, 2013.

[26] L. Khedher, J. Ram´ırez, J. M. G´orriz, A. Brahim, F. Segovia, A. s Disease Neuroimaging Initiative et al., "Early diagnosis of alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented mri images," Neurocomputing, vol. 151, pp. 139–150, 2015.

[27] J. Liu, J. Wang, "Improving Alzheimer's Disease Classification by Combining Multiple Measures," IEEE/ACM Transactions on Computational Biology and Bioinformatics, DOI 10.1109/TCBB.2017.

[28] H. I. Suk, C. Y. Wee, S. W. Lee, D. Shen, "State-space model with deep learning for functional dynamics estimation in resting-state fMRI," NeuroImage 129 (2016) 292–307.

[29] M. Liu, J. Zhang, E. Adeli, D. Shen "Landmark-based deep multi-instance learning for brain disease diagnosis." Medical Image Analysis 43 (2018) 157–168.

[30] R. Arma˜nanzas, M. Iglesias, D. A. Morales, and L. Alonso-Nanclares,"Voxel-Based Diagnosis of Alzheimer's Disease Using Classifier Ensembles," IEEE Journal of Biomedical and Health Informatics, VOL. 21, NO. 3, MAY 2017

[31] X. Zhu, H. I. Suk, D. Shen, "A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis," NeuroImage 100 (2014) 91–105

[32] J. Zhang, M. Liu, D. Shen, "Detecting Anatomical Landmarks from Limited Medical Imaging Data using Two-Stage Task-Oriented Deep Neural Networks." IEEE Transactions on Image Processing, DOI 10.1109/TIP.2017.2721106.