

Received October 20, 2020, accepted October 30, 2020, date of publication November 10, 2020, date of current version November 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3037107

Anatomical Landmarks and DAG Network Learning for Alzheimer's Disease Diagnosis

TIAN ZHU¹, CHONGFENG CAO², ZHISHUN WANG³, (Senior Member, IEEE), GUANGRUN XU⁴, AND JIANPING QIAO¹ 

¹Shandong Province Key Laboratory of Medical Physics and Image Processing Technology, School of Physics and Electronics, Shandong Normal University, Jinan 250014, China

²Department of Critical Care Medicine, Jinan Central Hospital, Cheeloo College of Medicine, Shandong University, Jinan 250012, China

³Department of Psychiatry, Columbia University, New York, NY 10027, USA

⁴Department of Neurology, Qilu Hospital of Shandong University, Jinan 250012, China

Corresponding author: Jianping Qiao (jpeqiao@sdu.edu.cn)


This work was supported in part by the National Natural Science Foundation of China under Grant 61603225, in part by the Natural Science Foundation of Shandong Province under Grant ZR2016FQ04, in part by the China Postdoctoral Science Foundation under Grant 2016M602182, in part by the Key Research and Development Foundation of Shandong Province under Grant 2016GGX101009, and in part by the Shandong Provincial Key Research and Development Plan under Grant 2017CXGC1504.

ABSTRACT The accurate diagnosis and prediction for individuals is crucial in computer-aided diagnosis of Alzheimer's disease (AD). The existing structural magnetic resonance imaging based classification methods of AD diagnosis mainly focus on the voxel level, region level and patch level morphological pattern analysis. However, most of these methods extract features with high dimension which may lead to overfitting problem. Besides, the interaction of different patches is not considered in the classifier ensemble. In this article, we propose a novel anatomical landmarks and directed acyclic graph (DAG) network feature learning based classification algorithm for the diagnosis of AD individuals. First, the anatomical feature patches of gray matter image are identified by the morphological and statistical analysis. Second, a simple and efficient DAG convolutional neural network is proposed to extract the discriminative deep features of image representation. Especially, the deep features are obtained by fusing feature maps of different network levels which contain semantic high-level and high-resolution low-level features. Finally, support vector machine and deep features are utilized to construct the classification model and predict the individual of AD. Experiments on three public datasets including ADNI-1, ADNI-2 and MIRIAD demonstrate that the proposed method can effectively improve the classification performance compared with the state-of-the-art methods for AD diagnosis.

INDEX TERMS Alzheimer's disease, structural magnetic resonance image, anatomical landmarks, directed acyclic graph network, classification.

I. INTRODUCTION

Alzheimer's disease (AD) is an increasingly prevalent neurodegenerative disease characterized by memory disorder, aphasia, disuse, blindness, executive dysfunction, personality and behavior changes [1], [2]. The AD pathogenesis is derived by the deposition of β -amyloid protein ($A\beta$) [3] and hyperphosphorylation of abnormal tau protein [4], leading to the formation of senile plaques, neuron apoptosis and eventually the decline of cognitive function of patients [5]. It accounts for 60% to 80% of dementia cases and has become the third leading cause of death after cancer and heart diseases

The associate editor coordinating the review of this manuscript and approving it for publication was Mauro Gaggero .

[6]. The AD patients also cause the financial and mental burden on their families. At present, the clinical diagnosis of AD mainly depends on the experience of the experts which is time-consuming and varies with individuals. Therefore, the computer-aided diagnosis of AD is necessary to assist the accurate diagnosis and treatment of the disease.

Various neuroimaging studies and computer-aided diagnosis methods have been actively applied to investigate the biological or neurological biomarkers of AD [7]–[11]. The accumulation level of biomarkers (i.e., $A\beta$ and Tau) [12], [13] are detected in cerebrospinal fluid (CSF) for AD diagnosis. The metabolic level of glucose in fluorodeoxyglucose positron emission tomography (FDG-PET) is measured as AD neuron dysfunction [14], [15]. As one of the noninvasive

imaging technologies, structural magnetic resonance imaging (sMRI) can sensitively capture the anatomical changes of the brain which has proved to be an effective tool for disease diagnosis [16], [17]. At present, the sMRI-based computer-aided methods for AD diagnosis are mainly divided into three categories: regions-of-interest (ROI) -based method, voxel-based method, and patch-based method [18]. In the ROI-based methods, the anatomic volume and thickness of predefined regions [19]–[24] are applied for AD diagnosis. Recently, the texture, shape and deep features of the hippocampal region are extracted by convolutional neural network (CNN) [25]–[27] to enhance the classification performance. However, the ROIs must be defined in advance and these regions may not cover all biomarkers or span multiple ROI, resulting in the loss of discriminant features. In the voxel-based methods [28]–[31] the brain microstructure changes are identified in voxel manner. The voxel-based morphology [32] is adopted to measure the density of gray matter (GM), white matter (WM) and CSF and find the abnormal brain regions between different groups. However, the high dimensional features in the voxel-based methods may cause the classifier to be challenged by overfitting. In the patch-based methods, a scale between voxel and region level is used to extract features with more information, which not only considers the subtle changes of brain tissue, but also reduces the feature dimension and avoids overfitting. For example, the discriminating anatomical landmarks are identified as image features by utilizing the morphological characteristics of voxels [33]–[36]. A deep learning framework ensemble based on landmarks is proposed for disease diagnosis [37]–[39]. Besides, a local patch-based weak classifier ensemble method is reported which combines multiple individual classifiers based on different subsets of local patches [40]. But the classification performance may be sub-optimal because different patches are respectively inputted into the individual classifiers which ignore the interaction of different patch features.

In this article, we propose a novel Alzheimer's disease classification method based on anatomical landmarks and directed acyclic graph network (DAG) feature learning (LDNFL). Compared with the traditional image representation and landmark based deep learning methods, the proposed method conducts a novel landmark based image representation scheme and deep features extraction based on a DAG CNN model. In the training stage, the anatomical feature patches of gray matter image are firstly identified by the morphological and statistical analysis. Then a DAG convolutional neural network is proposed to extract the discriminative deep features of image representation. In detail, feature maps in different network levels are fused which contain both strong semantic high-level and high-resolution low-level features. Finally, support vector machine (SVM) with deep fusion features is trained to obtain the discriminate model. In the testing stage, the deep features of the subjects in the testing dataset are extracted and inputted into SVM for individual AD diagnosis. Experimental results demonstrate that the proposed

method achieves better classification performance compared with the state-of-the-art methods.

The rest of the article is organized as follows. In Section II, we introduce the studied datasets and the preprocessing of sMRI data. The proposed anatomical landmarks and DAG network feature learning based classification algorithm is elaborated in Section III. Section IV describes the extensive experimental results and analyzes our work. We finally draw a conclusion in Section V.

II. MATERIALS

A. STUDIED DATASETS

Three public datasets are applied in this study including the Alzheimer's Disease Neuroimaging Initiative (ADNI-1), ADNI-2, and the Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD) dataset [38]. The subjects include 194 AD and 226 normal controls (NC) with 1.5T T1-weighted structure MRI data in the baseline ADNI-1 dataset. A total of 151 AD and 193 NC subjects with 3T T1-weighted structure MRI data are used in the baseline ADNI-2 dataset, in which the same subjects with ADNI-1 are excluded in order to ensure that all the data in this study are independent. The MIRIAD dataset contains 46 AD and 23 NC images. The demographic and clinical information of the studied subjects are presented in Table I.

B. DATA PREPROCESSING

The structural MRI data are preprocessed in a standard pipeline with CAT12 toolbox [41] which consists of four steps: 1) Bias correction. MRI is usually destroyed by a smooth and slowly changing bias field, resulting in inhomogeneous image intensity, thus affecting the accuracy of image automatic processing. Generally, there are two kinds of methods to measure and correct the intensity nonuniformity, including the methods that use parametric representations of image intensity distributions [42] and the methods that use non-parametric representations [43], [44]. In our method, an automatic bias field correction method [42] is used to measure and correct the inhomogeneous intensity in which the Mixture of Gaussians model with extra parameters that account for smooth intensity variations is applied. The number of the basis function will be small when the bias tends to be spatially smooth. 2) Tissue segmentation. The sMRI images are then segmented into WM, GM and CSF. 3) Space normalization. The images of all subjects are registered into the Montreal Neurological Institute (MNI) space by using the affine transformation and nonlinear registration for normalization. Specifically, the affine registration is applied for global geometric transformation of sMRI images by linear mapping. The nonlinear registration is used for local nonlinear deformation registration by adjusting different parts of the original image in different ways to achieve a more precise registration between images. 4) The modulation is finally performed by multiplying the jacobian of the deformation field to correct the volume changes of the segmented images.

TABLE 1. Demographic and clinical information of subjects in three public databases. Values are reported as mean±standard deviation (STD); MMSE: mini-mental state examination.

Dataset	Category	Number	Age (mean±std)	Gender (Male/Female)	MMSE (mean±std)
ADNI-1	AD	194	75.55±7.67	98/96	23.49±2.17
	NC	226	76.01±5.03	116/110	29.08±1.10
ADNI-2	AD	151	75.03±7.90	84/67	23.38±2.35
	NC	193	73.31±6.44	89/104	29.05±1.24
MIRIAD	AD	46	69.95±7.07	19/27	19.19±4.01
	NC	23	70.36±7.28	12/11	29.39±0.84

III. METHODS

The overview of the proposed anatomical landmarks and DAG network feature learning (LDNFL) based classification framework is illustrated in Fig. 1. It consists of four stages: 1) landmark location, 2) image representation, 3) deep features mining, 4) SVM classification. Specifically, the locations of landmarks are firstly determined by the statistics of the voxel morphological features of the preprocessed data in the training set. Second, image representation is obtained by feature patches selection. Third, we adopt DAG CNN model to dig the deep morphological features of image representation. Finally, the deep features are inputted into SVM to train the classification model. In the testing stage, the sMRI image of a new subject is preprocessed by the same standard pipeline with the training stage. After that, the patches of the located landmarks are inputted into the trained DAG network to obtain the deep features. Finally, the trained SVM classification model with the extracted deep features is applied for AD diagnosis.

A. LANDMARK LOCATION

We analyze and compare the voxel morphology of GM images between AD group and NC group in training set by using the two sample t-test [45] algorithm in the SPM12 toolbox to determine the locations of anatomical landmarks. Specifically, we first establish a two sample t-test statistical model and estimate the T-value of each voxel to generate a map of the statistical parameters. Then the significant abnormal areas are obtained under the predetermined contrast and significance level. The multiple comparison correction is performed due to the simultaneous testing on multiple voxels. If not, the probability of voxel misjudgment will be very high if the single test threshold is adopted. In this study,

family wise error (FWE) is applied for multiple comparison correction which divides the p value in the case of a single test by the number of comparisons to obtain a new p-value to judge the significance of the results. The significance level $p < 0.05$ (FWE correction) and clusters with the number of continuous voxels greater than 30 are used to identify the most significant discriminating regions between two groups, setting as $C = \{C_n\}_{n=1}^N$, where N is the number of discriminating regions. In each C_n , the locations with local minimum p-value that apart more than 8.0mm from each other are recognized as distinctive anatomical landmarks in MNI space. Finally, we transform the landmark locations from MNI space coordinates to the space voxel coordinates by using MRICroN medical image analysis software.

B. IMAGE REPRESENTATION

In order to extract regions with rich feature information for feature learning, we first extract anatomical feature patches of size $15 \times 15 \times 15$ from GM image with each anatomical landmark as the center. Then the Pearson correlation matrix between feature patches is calculated. Specifically, let $X_{m,n}$ be the average density value of all voxels in each feature patches of each subject, where $m = [1, 2, \dots, N_{sub}]$, $n = [1, 2, \dots, N_{patch}]$, N_{sub} and N_{patch} are the numbers of subjects and patches, respectively. The Pearson correlation coefficient between the two feature patches can be described by:

$$P(X_{:,n_1}, X_{:,n_2}) = \frac{\sum_{i=1}^{N_{sub}} (X_{i,n_1} - \bar{x}_{:,n_1})(X_{i,n_2} - \bar{x}_{:,n_2})}{\{\sum_{i=1}^{N_{sub}} (X_{i,n_1} - \bar{x}_{:,n_1})^2 \sum_{j=1}^{N_{sub}} (X_{j,n_2} - \bar{x}_{:,n_2})^2\}^{1/2}} \quad (1)$$

where n_1 and n_2 represent two different patches, respectively.

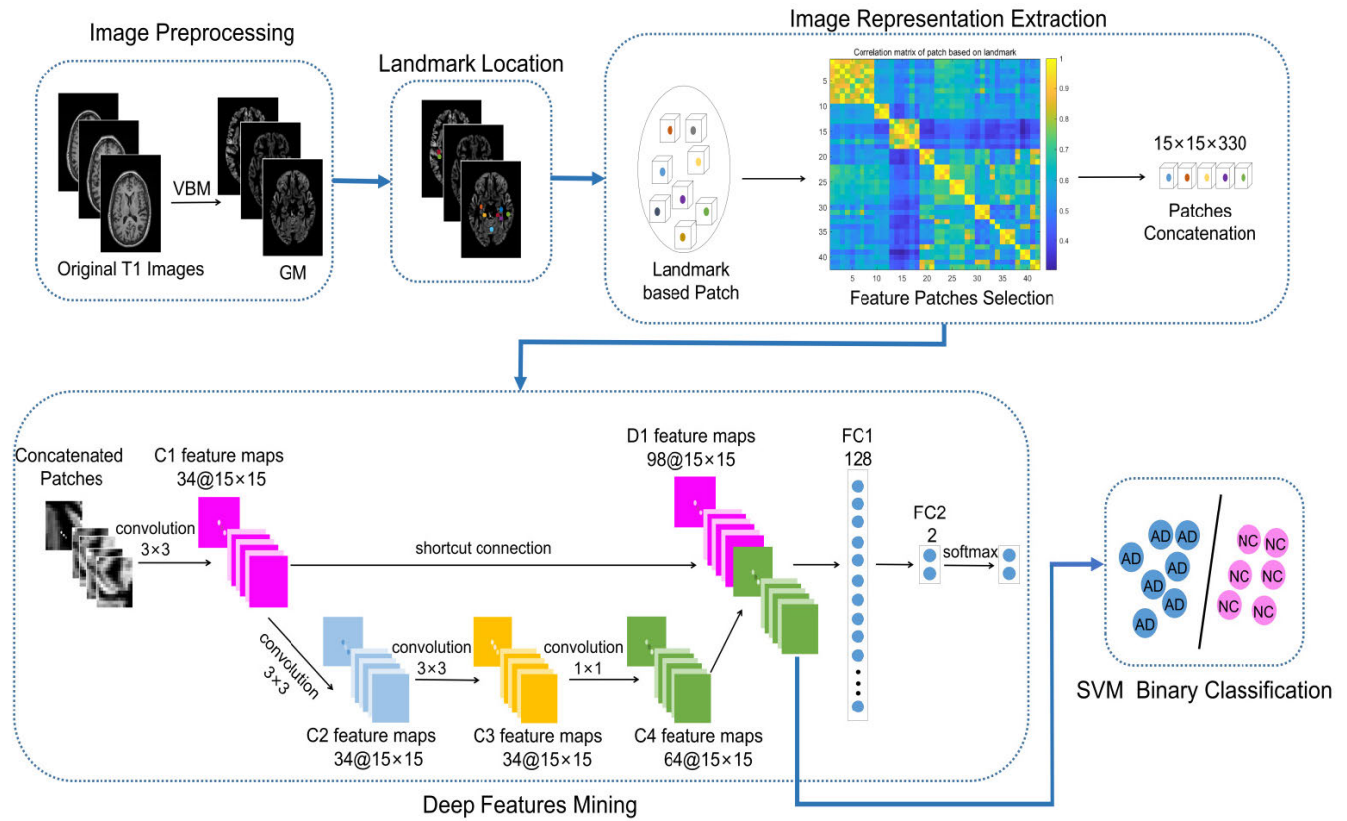


FIGURE 1. The framework of Alzheimer's disease classification algorithm based on anatomical landmarks and DAG network feature learning (LDNFL).

It is worth noting that patches that have less correlation with the patches in other clusters are excluded in order to ensure the diversity of features and avoid a large number of redundant features. Meanwhile, the patches with larger p-value landmarks are excluded to avoid the large area overlap between adjacent patches. We finally extract 22 patches based on landmarks in this study which are shown in Fig. 2. The structural changes of various regions in the brain are interrelated, we therefore concatenate the third dimension of feature patches as the image representation (with the size of $15 \times 15 \times 330$) which fuses the features of all landmarks.

C. DEEP FEATURES MINING

The convolutional neural network model exhibits high advantages in the aspect of feature extraction of high-dimensional data. Therefore, we propose a novel DAG convolutional neural network to extract the discriminative deep fusion features of image representation in a supervised manner, which can alleviate the network degradation and integrate the features of different levels. The schematic diagram of the network model in Fig. 1 shows that the DAG CNN consists of four convolutional layers and two fully connected layers. Specifically, the image represents are fed into CNN with an input layer of size $15 \times 15 \times 330$. Then input images are convoluted successively by three sequential convolutional layers (i.e., C1, C2, C3, C4) and activated by a leaky rectified linear unit (ReLU) after each convolution. The kernel sizes of C1,

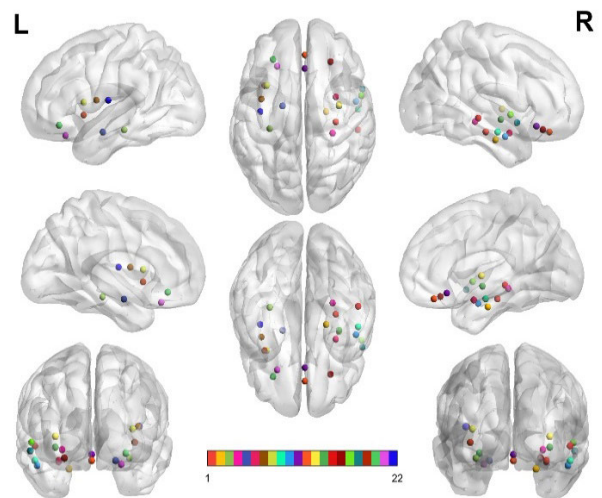


FIGURE 2. The selected twenty-two anatomical landmarks.

C2, C3, C4 layers are $3 \times 3, 3 \times 3, 3 \times 3, 1 \times 1$, respectively. Convolution with kernel of size 1×1 can increase the non-linearity of the network and make the network express more complex features. Then, the output of C1 is concatenated with the output of C4 in depth by shortcut connection (i.e., skip connection) to form the proposed DAG network. The extracted deep features contain both semantic high-level and high-resolution low-level features. The output of the depth

concatenation layer (D1) is sent to two successive fully connected layers (i.e., FC1, FC2) and the softmax output layer in sequence.

Assume that the h -th layer of the network is a convolutional layer, each feature map of the h -th layer can be described by:

$$y_k^h = g \left(\sum_{i \in M_k} y_i^{h-1} * w_{ik}^h + b_k^h \right) \quad (2)$$

where w represents the weight, b is the bias, M_k represents the set of the input feature maps and g indicates the activation function. Specifically, the feature maps of the $(h-1)$ -th layer are firstly convoluted with the convolution kernel. Then the feature maps of the h -th layer are formed by performing the activation function. Each feature map y_k^h ($k = 1, 2, 3, \dots, N_k$) is obtained by combining the same positions of the convolutional results of the input maps, where N_k is the number of the learnable kernels.

The CNN model is optimized by Adaptive moment estimation (Adam) algorithm [46], which designs independent adaptive learning rates for different parameters by calculating the first and second moment estimations of the gradient. The first moment is estimated by:

$$m_n = \alpha_1 m_{n-1} + (1 - \alpha_1) \nabla E(\theta_n) \quad (3)$$

where n is the iteration number, θ indicates the parameter vector, $E(\theta)$ is the loss function, $\nabla E(\theta)$ represents the gradient of the loss function and α_1 means the exponential decay rate of the first moment estimation. The second moment is estimated by:

$$v_n = \alpha_2 v_{n-1} + (1 - \alpha_2) [\nabla E(\theta_n)]^2 \quad (4)$$

where α_2 is the exponential decay rate of the second moment estimation.

It is necessary to correct the bias of m_n and v_n to reduce the influence of the initial values. The bias-corrected first and second moment estimations can be recalculated by:

$$\hat{m}_n = m_n / (1 - \alpha_1^n) \quad (5)$$

$$\hat{v}_n = v_n / (1 - \alpha_2^n) \quad (6)$$

Finally, the two moving averages are used to adaptively update the network parameters by:

$$\theta_{n+1} = \theta_n - \alpha \bullet \hat{m}_n / (\sqrt{\hat{v}_n} + \varepsilon) \quad (7)$$

where α represents the learning rate, ε is set to be 10^{-8} which avoids a divisor of zero.

The exponential decay rates (i.e., α_1 and α_2) of the first and second moment estimations are 0.8 and 0.999, respectively. We train CNN for 18 epochs with an initial learning rate (i.e., α) of 0.003 and update the weights in mini-batches size of 64 per batch. The number of epochs for dropping the learning rate is 5. The L2 regularization coefficient is set to be 0.001 which reduces the complexity and instability of the model and avoids the risk of over fitting.

D. SVM CLASSIFICATION

SVM is a supervised classifier which finds an optimal hyper-plane to maximize the margin between the classes. The deep features of each training subject are extracted from the D1 layer of the DAG CNN model in the form of a row vector with the size of 1×22050 . The SVM classifier with linear kernel is constructed based on the deep features of the training images. In the real time test scenario, the sMRI image of a new testing subject is preprocessed and normalized to a standard MNI template. The location of the 22 landmarks that are determined during the training stage are used for the normalized testing data directly. After that, the deep features of the 22 landmarks patches are extracted by using the trained DAG network and then inputted into the SVM classification model to determine whether this subject has developed AD or not.

IV. EXPERIMENTS

A. METHODS FOR COMPARISON

1) ROI-based method [47]. We adopt the anatomical automatic labeling (AAL) [48] that is registered to MNI space in CAT12 toolbox to divide the GM image into 90 ROIs. Then gray matter volumes of 90 ROIs are extracted as ROI features (with the size of 1×90) of each subject. Finally, the extracted ROI features are used for disease classification via SVM.

2) Voxel-based morphometry (VBM) method [32]. First, GM images generated by normalization, tissue segmentation and modulation are spatially smoothed with 8mm full-width-half-maximum (FWHM) gaussian kernel to improve the signal-to-noise ratio of images. Then, the statistical model of each voxel in GM is established and the brain regions with significant differences between the two groups are obtained. The abnormal brain tissue density is extracted as the feature representation of MR images. Finally, these features on voxel level are inputted into SVM for classification.

3) Landmark based feature representation (LFR) method [33]. The high feature dimension generated by VBM may lead to the high complexity and overfitting of classification model. Therefore, the tissue density around anatomical landmark is selected as the image feature representation in voxel manner in order to reduce dimension. Specifically, GM patch of size $15 \times 15 \times 15$ is extracted with the landmark as the center and arranged into a row vector. Then, all row vectors of each subject are concatenated into a longer vector as the feature representation of MRI to be fed into SVM classifier for AD classification.

4) Patch-based deep features concatenation (PDFC) method [38]. The method trains multiple CNN models to learn deep features of patches. It is worth noting that each CNN model corresponds to a landmark based patch. In the experiment, a total of 22 CNN models are trained. Then, the features of depth concatenation layer (D1) in each network model are extracted in the form of row vectors and concatenated into a longer vector which is sent to SVM with linear kernel for AD diagnosis.

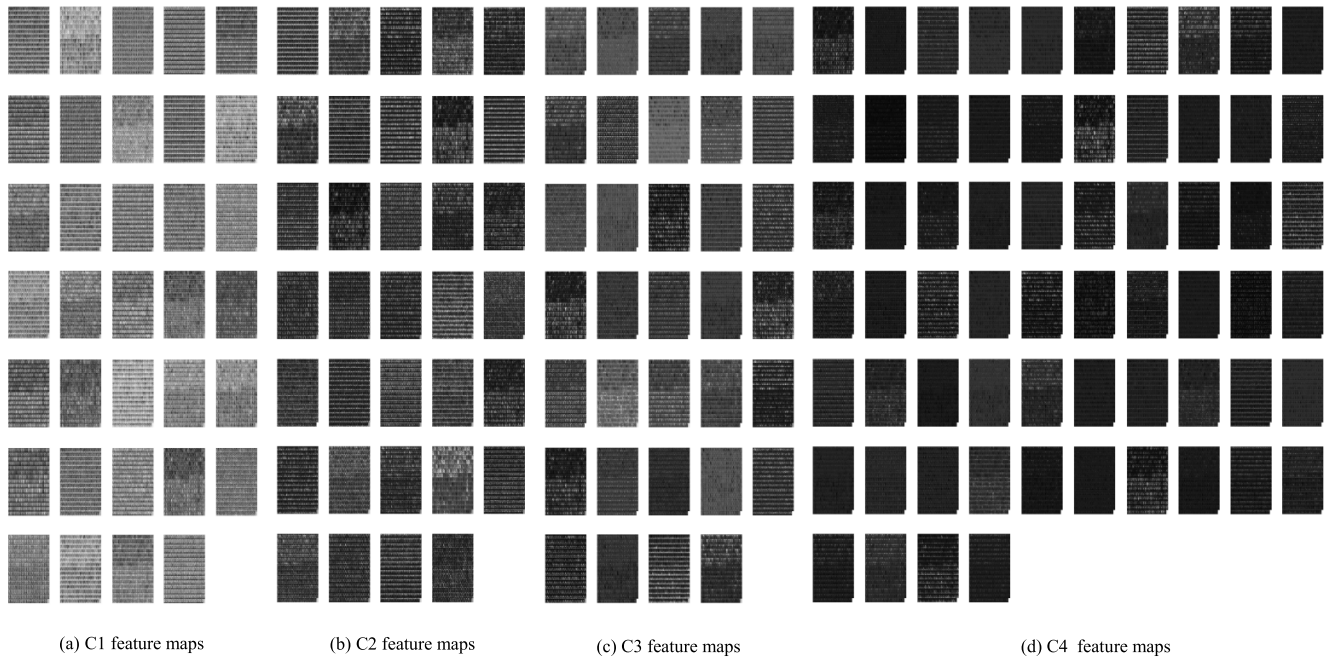


FIGURE 3. Feature maps of ADNI-2 database extracted by convolutional layer of DAG CNN model.

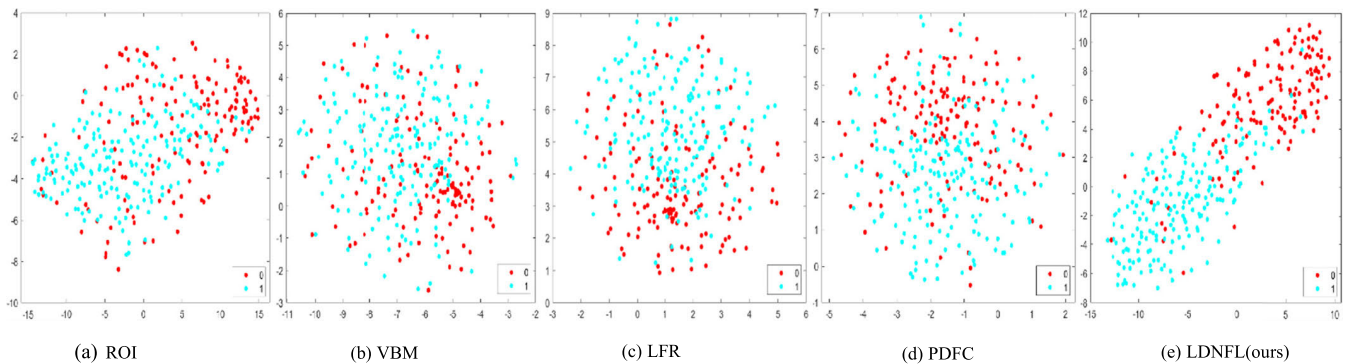


FIGURE 4. Visualization of the extracted features in different methods.

5) Image representation based on patches concatenation. For each subject, we extract 22 GM patches based on landmarks. Then all patches are concatenated as image representation to mine features by training the proposed DAG CNN models in a supervised manner. The next classification mainly includes two methods. One is fully connected layer and softmax layer for classification (denoted as LDNC). The learned features are mapped to the sample label space by full connection layer and classification probability is predicted through the softmax layer. The other is SVM classification. Features from D1 layer (i.e., our proposed LDNFL method) are fed to SVM classifier for disease classification.

B. EXPERIMENTAL SETTINGS

In our experiment, there are eight evaluation indices for evaluating the performance of disease classification including:

1) accuracy (ACC), 2) sensitivity (SEN): the probability of correct diagnosis in patients, 3) specificity (SPE): the

probability of correct diagnosis in normal subjects, 4) precision (P): $P = TP / (TP + FP)$, the probability of actual disease in the population judged to be diseased, 5) F-Measure [49]: $F = 2 \times P \times R / (P + R)$, which is used for comprehensive evaluation, 6) receiver operating characteristic curve (ROC), 7) area under ROC (AUC), 8) precise-recall curve (P-R) [50], where TP denotes true positive, TN denotes true negative, FP denotes false positive, FN denotes false negative.

C. EXPERIMENTAL RESULT

In order to verify the effectiveness of the proposed method, the ADNI-1 dataset is used as the training set while ADNI-2 and MIRIAD datasets are used as the testing set, respectively. The selected 22 landmarks shown in Fig. 2 mainly locate in bilateral hippocampus, bilateral parahippocampal gyrus, bilateral amygdala, bilateral insula and bilateral fusiform gyrus, which have been proved to be important markers to

TABLE 2. Classification results of AD / NC (the classification model is trained on ADNI-1 and tested on ADNI-2).

<i>Methods</i>	<i>ACC</i>	<i>SEN</i>	<i>SPE</i>	<i>P</i>	<i>F-Measure</i>	<i>AUC</i>
ROI	84.59%	80.13%	88.08%	84.03%	82.03%	90.73%
VBM	87.21%	74.83%	96.89%	94.96%	83.70%	94.87%
LFR	87.79%	78.15%	95.34%	92.91%	84.89%	94.51%
PDFC	88.95%	80.13%	95.85%	93.80%	86.43%	95.14%
LDNC	89.83%	83.44%	94.82%	92.65%	87.80%	95.15%
LDNFL(ours)	91.57%	87.42%	94.82%	92.96%	90.10%	95.77%

distinguish AD from NC in the previous studies [51]–[53]. Deep features are extracted by the proposed DAG network. The feature maps of different convolutional layers are shown in Fig. 3. Fig. 3 (a), (b) and (c) show 34 feature maps in C1, C2 and C3 convolution layers, respectively. Fig. 3 (d) shows 64 feature maps in C4 convolution layer. There are 344 (the number of the testing subjects) blocks with the size of 15×15 in each feature map. The features of C1, C4 and D1 layers are extracted and inputted into SVM for classification. The classification accuracy is 90.99%, 89.53% and 91.57%, respectively. Therefore, the classification accuracy is improved by increasing the depth of CNN model.

The t-distributed stochastic neighbor embedding (t-SNE) dimension reduction algorithm [54] is applied to reduce the dimension of features that are inputted into SVM for classification from high dimension to two dimension in order to visualize the discriminant ability of extracted features, as shown in Fig. 4. The features extracted by the proposed feature extraction method have better separability compared with other methods. The classification comparison between the proposed method and the traditional methods performed on ADNI-2 testing dataset is shown in Table II. It can be seen that the proposed LDNFL method has the highest accuracy, F-measure and AUC values compared with other feature methods. Moreover, the proposed method fuses different patch features into a single CNN model which achieves better classification performance compared with isolated multiple neural network models. Furthermore, our LDNFL method combines CNN and SVM which shows better classification results than that only uses SVM (i.e., LFR). We also implement classification by using the softmax layer (the LDNC method), the accuracy is 89.83%. As shown in Table II, the accuracy, F-Measure and AUC of our proposed method are improved by 1.74%, 2.3% and 0.62% compared with that of the LDNC method, respectively. The framework

combined with CNN feature extraction and SVM classifier achieves better results than the softmax layer classifier in some previous studies [55]–[59], which is consistent with our findings. In addition, we implement classifiers with less time complexity such as linear discriminant analysis with the same deep features. The accuracy, F-Measure and AUC of linear discriminant analysis classifier are 88.37%, 85.72% and 94.89%, respectively, which are less than that of SVM results (91.57%, 90.10%, 95.77%), which is consistent with the previous studies [60], [61]. The classification accuracy is determined by both features and classification model. When feature quality is really good, different classification model may obtain different accuracy because the features may be mapped to different spaces by linear or non-linear functions.

Therefore, the DAG network is used for deep feature extraction and SVM is applied for classification in the proposed method which achieves the better accuracy results compared with other methods. The ROC and P-R curves of the proposed method and other comparison methods are plotted in Fig. 5. The proposed LDNFL framework also reveals the best performance in ROC and P-R curves.

We further perform different methods with ADNI-1 as the training set and MIRIAD as the testing set. The classification performance of our framework is exhibited in Table III and Fig. 6. Similarly, our proposed method is generally superior to other existing methods in AD and NC classification.

Table IV shows the performance comparison of the proposed method with the state-of-the-art methods. It can be seen that the proposed method achieves better diagnosis results compared with the three kinds of methods including ROI+CNN model [25], ensemble classifier [38], [39], and end-to-end hierarchical convolution network learning based on whole brain image patches [18]. In addition, larger datasets are applied in the proposed method which improves the generalization and effectiveness of the classification model.

TABLE 3. Classification results of AD / NC (the classification model is trained on ADNI-1 and tested on MIRIAD)

Methods	ACC	SEN	SPE	P	F-Measure	AUC
ROI	89.86%	89.13%	91.30%	95.35%	92.14%	96.98%
VBM	91.30%	93.48%	86.96%	93.48%	93.48%	97.26%
LFR	89.86%	84.78%	100.0%	100.0%	91.76%	97.45%
PDFC	91.30%	89.13%	95.65%	97.62%	93.18%	97.16%
LDNC	91.30%	86.96%	100.0%	100.0%	93.03%	96.79%
LDNFL(ours)	92.75%	91.30%	95.65%	97.67%	94.38%	97.64%

TABLE 4. Performance comparison of the proposed method with the state-of-the-art methods

Reference	Subject	ACC	SEN	SPE	F-Measure	AUC
Lian et al [18]	ADNI-1,ADNI-2	90.30%	82.40%	96.50%	-	95.10%
Cui et al [25]	192AD,223NC	92.29%	90.63%	93.72%	-	96.95%
Zhang et al [33]	ADNI-1	83.10%	80.50%	85.10%	-	-
Liu et al [38]	ADNI-1,ADNI-2	90.56%	87.42%	93.03%	89.10%	95.74%
Ahmed et al [39]	77 AD, 129 NC	85.55%	-	-	-	-
LDNFL(ours)	ADNI-1,ADNI-2	91.57%	87.42%	94.82%	90.10%	95.77%

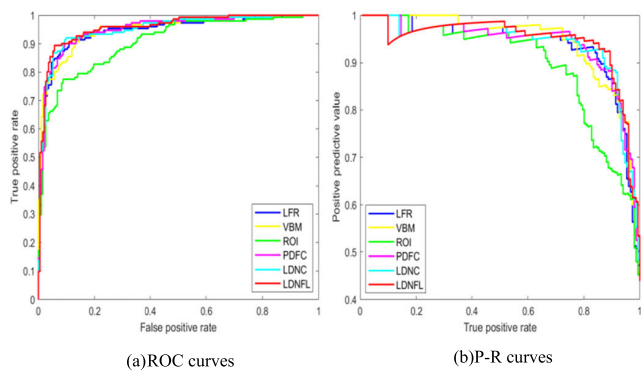


FIGURE 5. ROC and P-R curves of different methods for AD classification (trained on ADNI-1 and tested on ADNI-2).

Therefore, the proposed LDNFL method can automatically identify the anatomical feature patches and extract

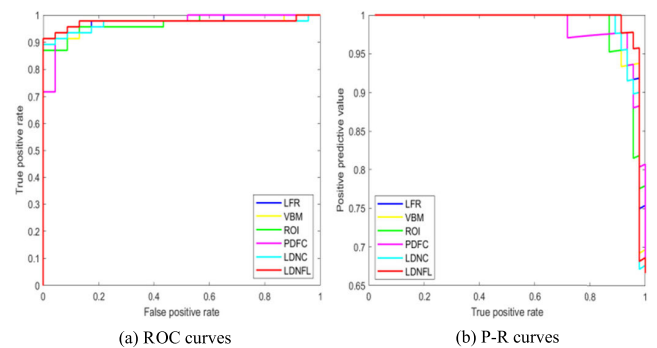


FIGURE 6. ROC and P-R curves of different methods for AD classification (trained on ADNI-1 and tested on MIRIAD).

the deep features with strong semantics and high resolution through training a unified DAG CNN model. Specifically, the proposed method can capture the subtle abnormal

structural changes of feature patches without defining ROIs in advance compared with the ROI-based methods. In addition, our landmark-based image representation method reduces the feature dimension and avoids the risk of overfitting compared with the VBM-based methods. Moreover, all landmarks based anatomical feature patches are connected and inputted into a single CNN model for feature learning. In this way, compared with the methods that construct isolated CNN models for each landmark, the proposed method fuses all landmarks features, considering the mutual influence of all features. Besides, deep features with strong semantic and high resolution are obtained by performing skip connection in DAG network which also enhances classification performance.

V. CONCLUSION

In this article, we propose an anatomical landmarks and DAG network feature learning method for AD diagnosis. The GM image patches based on landmarks are extracted and concatenated as image feature representations. Then a DAG convolutional neural network is applied to automatically extract the discriminative deep features of image representation. The SVM with low dimensional deep features is finally used to construct the classification model for AD diagnosis. The proposed method achieves superior classification and prediction performances over other state-of-the-art methods on three popular public datasets.

In the future work, we will focus on the model generalization in real world conditions since the landmark location in the proposed method is isolated from the DAG network feature learning, which may lead to sub-optimal performance [38]. Furthermore, multiclass neuroimaging classification instead of binary classification will be developed for the automatic diagnosis of different brain diseases.

REFERENCES

- [1] S. P. Cass, "Alzheimer's disease and exercise: A literature review," *Current Sports Med. Rep.*, vol. 16, no. 1, pp. 19–22, Jan. 2017.
- [2] T. Jonsson *et al.*, "A mutation in APP protects against Alzheimer's disease and age-related cognitive decline," *Nature*, vol. 488, no. 7409, pp. 96–99, Aug. 2012.
- [3] M. LaFerla, K. N. Green, and S. Oddo, "Intracellular amyloid- β in Alzheimer's disease," *Nature Rev. Neurosci.*, vol. 8, no. 7, pp. 499–509, 2007.
- [4] K. Blennow, E. Vanmechelen, and H. Hampel, "CSF total tau, A β 42 and phosphorylated tau protein as biomarkers for Alzheimer's disease," *Mol. Neurobiol.*, vol. 24, nos. 1–3, pp. 87–98, 2001.
- [5] J. Hardy and D. J. Selkoe, "The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics," *Science*, vol. 297, no. 5580, pp. 353–356, Jul. 2002.
- [6] B. D. James, S. E. Leurgans, L. E. Hebert, P. A. Scherr, K. Yaffe, and D. A. Bennett, "Contribution of alzheimer disease to mortality in the united states," *Neurology*, vol. 82, no. 12, pp. 1045–1050, Mar. 2014.
- [7] G. Chen, B. D. Ward, C. Xie, W. Li, Z. Wu, J. L. Jones, M. Franczak, P. Antuono, and S.-J. Li, "Classification of alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional MR imaging," *Radiology*, vol. 259, no. 1, pp. 213–221, Apr. 2011.
- [8] J. Qiao, Y. Lv, C. Cao, Z. Wang, and A. Li, "Multivariate deep learning classification of Alzheimer's disease based on hierarchical partner matching independent component analysis," *Frontiers Aging Neurosci.*, vol. 10, Dec. 2018, Art. no. 417.
- [9] Y. Shi, H.-I. Suk, Y. Gao, S.-W. Lee, and D. Shen, "Leveraging coupled interaction for multimodal Alzheimer's disease diagnosis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 1, pp. 186–200, Jan. 2020.
- [10] M. S. Tahaei, M. Jalili, and M. G. Knyazeva, "Synchronizability of EEG-based functional networks in early Alzheimer's disease," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 5, pp. 636–641, Sep. 2012.
- [11] A. Khazaei, A. Ebrahimpour, and A. Babajani-Feremi, "Classification of patients with MCI and AD from healthy controls using directed graph measures of resting-state fMRI," *Behavioural Brain Res.*, vol. 322, pp. 339–350, Mar. 2017.
- [12] J. Fortea, M. Carmona-Iragui, B. Benezam, S. Fernández, L. Videla, I. Barroeta, D. Alcolea, J. Pegueroles, L. Muñoz, O. Belbin, M. J. de Leon, A. M. Maceski, C. Hirtz, J. Clarimón, S. Videla, C. Delaby, S. Lehmann, R. Blesa, and A. Lleó, "Plasma and CSF biomarkers for the diagnosis of Alzheimer's disease in adults with down syndrome: A cross-sectional study," *Lancet Neurol.*, vol. 17, no. 10, pp. 860–869, Oct. 2018.
- [13] J. L. Molinuevo, J. D. Gispert, B. Dubois, M. T. Heneka, A. Lleó, S. Engelborghs, J. Pujol, L. C. de Souza, D. Alcolea, F. Jessen, M. Sarazin, F. Lamari, M. Balasa, A. Antonell, and L. Rami, "The AD-CSF-Index discriminates Alzheimer's disease patients from healthy controls: A validation study," *J. Alzheimer's Disease*, vol. 36, no. 1, pp. 67–77, Jun. 2013.
- [14] Y. Ding *et al.*, "A deep learning model to predict a diagnosis of alzheimer disease by using ^{18}F -FDG PET of the brain," *Radiology*, vol. 290, no. 2, pp. 456–464, Feb. 2019.
- [15] L. Mosconi, R. Mistur, R. Switalski, W. H. Tsui, L. Glodzik, Y. Li, E. Pirraglia, S. De Santi, B. Reisberg, T. Wisniewski, and M. J. de Leon, "FDG-PET changes in brain glucose metabolism from normal cognition to pathologically verified Alzheimer's disease," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 36, no. 5, pp. 811–822, May 2009.
- [16] G. B. Frisoni, N. C. Fox, C. R. Jack, P. Scheltens, and P. M. Thompson, "The clinical use of structural MRI in alzheimer disease," *Nature Rev. Neurol.*, vol. 6, no. 2, pp. 67–77, Feb. 2010.
- [17] J. Koikkalainen, H. Rhodius-Meester, A. Tolonen, F. Barkhof, B. Tijms, A. W. Lemstra, T. Tong, R. Guerrero, A. Schuh, C. Ledig, D. Rueckert, H. Soininen, A. M. Remes, G. Waldemar, S. Hasselbalch, P. Mecocci, W. van der Flier, and J. Lötjönen, "Differential diagnosis of neurodegenerative diseases using structural MRI data," *NeuroImage, Clin.*, vol. 11, pp. 435–449, Jan. 2016.
- [18] C. Lian, M. Liu, J. Zhang, and D. Shen, "Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 880–893, Apr. 2020.
- [19] J. Lötjönen, R. Wolz, J. Koikkalainen, V. Julkunen, L. Thurfjell, R. Lundqvist, G. Waldemar, H. Soininen, and D. Rueckert, "Fast and robust extraction of hippocampus from MR images for diagnostics of Alzheimer's disease," *NeuroImage*, vol. 56, no. 1, pp. 185–196, May 2011.
- [20] C. R. Jack *et al.*, "Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment," *Neurology*, vol. 52, no. 7, pp. 1397–1403, Apr. 1999.
- [21] C. Platero and M. C. Tobar, "A fast approach for hippocampal segmentation from T1-MRI for predicting progression in Alzheimer's disease from elderly controls," *J. Neurosci. Methods*, vol. 270, pp. 61–75, Sep. 2016.
- [22] B. Fischl and A. M. Dale, "Measuring the thickness of the human cerebral cortex from magnetic resonance images," *Proc. Nat. Acad. Sci. USA*, vol. 97, no. 20, pp. 11050–11055, Sep. 2000.
- [23] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehericy, M.-O. Habert, M. Chupin, H. Benali, and O. Colliot, "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database," *NeuroImage*, vol. 56, no. 2, pp. 766–781, May 2011.
- [24] P. P. D. M. Oliveira, R. Nitrini, G. Busatto, C. Buchpiguel, J. R. Sato, and E. Amaro, "Use of SVM methods with surface-based cortical and volumetric subcortical measurements to detect Alzheimer's disease," *J. Alzheimer's Disease*, vol. 19, no. 4, pp. 1263–1272, Mar. 2010.
- [25] R. Cui and M. Liu, "Hippocampus analysis by combination of 3-D DenseNet and shapes for Alzheimer's disease diagnosis," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 5, pp. 2099–2107, Sep. 2019.
- [26] F. Li and M. Liu, "A hybrid convolutional and recurrent neural network for hippocampus analysis in Alzheimer's disease," *J. Neurosci. Methods*, vol. 323, pp. 108–118, Jul. 2019.
- [27] K. Zhao *et al.*, "Independent and reproducible hippocampal radiomic biomarkers for multisite Alzheimer's disease: Diagnosis, longitudinal progress and biological basis," *Sci. Bull.*, vol. 65, no. 13, pp. 1103–1113, 2020.

- [28] C. Tas, H. Mogulkoc, G. Eryilmaz, I. Gogcegoz-Gul, T. T. Erguzel, B. Metin, and N. K. Tarhan, "Discriminating schizophrenia and schizo-obsessive disorder: A structural MRI study combining VBM and machine learning methods," *Neural Comput. Appl.*, vol. 29, no. 2, pp. 377–387, Jan. 2018.
- [29] G. F. Busatto, B. S. Diniz, and M. V. Zanetti, "Voxel-based morphometry in Alzheimer's disease," *Expert Rev. Neurotherapeutics*, vol. 8, no. 11, pp. 1691–1702, Nov. 2008.
- [30] C. Möller, Y. A. L. Pijnenburg, W. M. van der Flier, A. Versteeg, B. Tijms, J. C. de Munck, A. Hafkemeijer, S. A. R. B. Rombouts, J. van der Grond, J. van Swieten, E. Dopfer, P. Scheltens, F. Barkhof, H. Vrenken, and A. M. Wink, "Alzheimer disease and behavioral variant frontotemporal dementia: Automatic classification based on cortical atrophy for single-subject diagnosis," *Radiology*, vol. 279, no. 3, pp. 838–848, Jun. 2016.
- [31] J. Zhang, B. Yan, X. Huang, P. Yang, and C. Huang, "The diagnosis of Alzheimer's disease based on voxel-based morphometry and support vector machine," in *Proc. 4th Int. Conf. Natural Comput.*, Shandong, China, 2008, pp. 197–201.
- [32] J. Ashburner and K. J. Friston, "Voxel-based morphometry—The methods," *NeuroImage*, vol. 11, no. 6, pp. 805–821, Jun. 2000.
- [33] J. Zhang, Y. Gao, Y. Gao, B. C. Munsell, and D. Shen, "Detecting anatomical landmarks for fast Alzheimer's disease diagnosis," *IEEE Trans. Med. Imag.*, vol. 35, no. 12, pp. 2524–2533, Dec. 2016.
- [34] X. Zhu, K. Thung, J. Zhang, and D. Shen, "Fast neuroimaging-based retrieval for Alzheimer's disease analysis," in *Proc. MLMI*, Athens, Greece, 2016, pp. 313–321.
- [35] J. Zhang, M. Liu, L. An, Y. Gao, and D. Shen, "Alzheimer's disease diagnosis using landmark-based features from longitudinal structural MR images," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 6, pp. 1607–1616, Nov. 2017.
- [36] J. Zhang, M. Liu, L. An, Y. Gao, and D. Shen, "Landmark-based Alzheimer's disease diagnosis using longitudinal structural MR images," in *Proc. MICCAI*, Athens, Greece, 2016, pp. 35–45.
- [37] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Landmark-based deep multi-instance learning for brain disease diagnosis," *Med. Image Anal.*, vol. 43, pp. 157–168, Jan. 2018.
- [38] M. Liu, J. Zhang, D. Nie, P.-T. Yap, and D. Shen, "Anatomical landmark based deep feature representation for MR images in brain disease diagnosis," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 5, pp. 1476–1485, Sep. 2018.
- [39] S. Ahmed, K. Y. Choi, J. J. Lee, B. C. Kim, G.-R. Kwon, K. H. Lee, and H. Y. Jung, "Ensembles of patch-based classifiers for diagnosis of Alzheimer diseases," *IEEE Access*, vol. 7, pp. 73373–73383, 2019.
- [40] M. Liu, D. Zhang, and D. Shen, "Ensemble sparse classification of Alzheimer's disease," *NeuroImage*, vol. 60, no. 2, pp. 1106–1116, Apr. 2012.
- [41] F. Farokhian, I. Beheshti, D. Sone, and H. Matsuda, "Comparing CAT12 and VBM8 for detecting brain morphological abnormalities in temporal lobe epilepsy," *Frontiers Neurol.*, vol. 8, p. 428, Aug. 2017.
- [42] J. Ashburner and K. J. Friston, "Unified segmentation," *NeuroImage*, vol. 26, no. 3, pp. 839–851, Jul. 2005.
- [43] T. Imtiaz, S. Rifat, S. A. Fattah, and K. A. Wahid, "Automated brain tumor segmentation based on multi-planar superpixel level features extracted from 3D MR images," *IEEE Access*, vol. 8, pp. 25335–25349, Dec. 2020.
- [44] L. G. Nyul, J. K. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization," *IEEE Trans. Med. Imag.*, vol. 19, no. 2, pp. 143–150, Feb. 2000.
- [45] N. A. C. Cressie and H. J. Whitford, "How to use the two sample t-Test," *Biometrical J.*, vol. 28, no. 2, pp. 131–148, 1986.
- [46] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *Proc. ICLR*, San Diego, CA, USA, 2015, pp. 1–41.
- [47] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *NeuroImage*, vol. 55, no. 3, pp. 856–867, 2011.
- [48] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "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, Jan. 2002.
- [49] N. Chinchor and B. Sundheim, "MUC-5 evaluation metrics," in *Proc. 5th Conf. Message Understand. (MUC)*, Baltimore, MD, USA, 1993, pp. 69–78.
- [50] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn. (ICML)*, New York, NY, USA, 2006, pp. 233–240.
- [51] A. Basher, K. Y. Choi, J. J. Lee, B. Lee, B. C. Kim, K. H. Lee, and H. Y. Jung, "Hippocampus localization using a two-stage ensemble Hough convolutional neural network," *IEEE Access*, vol. 7, pp. 73436–73447, 2019.
- [52] A. Chaddad, C. Desrosiers, and T. Niazi, "Deep radiomic analysis of MRI related to Alzheimer's disease," *IEEE Access*, vol. 6, pp. 58213–58221, 2018.
- [53] M. Atiya, B. T. Hyman, M. S. Albert, and R. Killiany, "Structural magnetic resonance imaging in established and prodromal Alzheimer disease: A review," *Alzheimer Disease Associated Disorders*, vol. 17, no. 3, pp. 177–195, Jul. 2003.
- [54] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
- [55] P. Chagas, L. Souza, I. Araújo, N. Aldeman, A. Duarte, M. Angelo, W. L. C. dos-Santos, and L. Oliveira, "Classification of glomerular hypercellularity using convolutional features and support vector machine," *Artif. Intell. Med.*, vol. 103, Mar. 2020, Art. no. 101808.
- [56] D.-X. Xue, R. Zhang, H. Feng, and Y.-L. Wang, "CNN-SVM for microvascular morphological type recognition with data augmentation," *J. Med. Biol. Eng.*, vol. 36, no. 6, pp. 755–764, Dec. 2016.
- [57] Y. Peng, J. Cai, T. Wu, G. Cao, N. Kwok, S. Zhou, and Z. Peng, "A hybrid convolutional neural network for intelligent wear particle classification," *Tribol. Int.*, vol. 138, pp. 166–173, Oct. 2019.
- [58] X. Sun, L. Liu, C. Li, J. Yin, J. Zhao, and W. Si, "Classification for remote sensing data with improved CNN-SVM method," *IEEE Access*, vol. 7, pp. 164507–164516, Nov. 2019.
- [59] X.-X. Niu and C. Y. Suen, "A novel hybrid CNN-SVM classifier for recognizing handwritten digits," *Pattern Recognit.*, vol. 45, no. 4, pp. 1318–1325, Apr. 2012.
- [60] Z. Sun, G. Bebis, X. Yuan, and S. J. Louis, "Genetic feature subset selection for gender classification: A comparison study," in *Proc. 6th IEEE Workshop Appl. Comput. Vis. (WACV)*, Orlando, FL, USA, Dec. 2002, pp. 165–170.
- [61] N. Gerhardt, S. Schwolow, S. Rohn, P. R. Pérez-Cacho, H. Galán-Soldevilla, L. Arce, and P. Weller, "Quality assessment of olive oils based on temperature-ramped HS-GC-IMS and sensory evaluation: Comparison of different processing approaches by LDA, kNN, and SVM," *Food Chem.*, vol. 278, pp. 720–728, Apr. 2019.



TIAN ZHU received the B.S. degree in electronic information engineering from Shandong Normal University, in 2017, where she is currently pursuing the M.S. degree with the School of Physics and Electronics. Her research interests include deep learning and MRI image processing and classification.



CHONGFENG CAO received the clinical degree in medical master from Shandong University, Jinan, China, in 2013.

He is currently an Associate Chief Physician in critical care with Jinan Central Hospital, Jinan. His current research interests include myocardial hypertrophy and nervous system dysfunction of severe sepsis.



ZHISHUN WANG (Senior Member, IEEE) received the Ph.D. degree in signal processing from Southeast University in 1997.

He is currently an Associate Professor in brain imaging with the Department of Psychiatry, Columbia University, and the New York State Psychiatric Institute, where he is leading a group to conduct pioneered research to discover brain circuits that govern human cognitive process and psychiatric disorders. His research interests include functional and structural MRI, multidimensional signal/image processing, neural networks, machine learning, chaos and fractal in the brain and the gut, independent component analysis, wavelet, and functional and effective brain connectivity. He is a Senior Member of the IEEE Signal Processing Society. His doctoral thesis was granted “Excellent Doctoral Thesis Award in Year 2000” from the Academic Committee of the State Council of China.



GUANGRUN XU received the M.D. degree in neurology from the Tongji Medical College, Huazhong University of Science and Technology, in 2000.

He is currently the Chief Physician with the Department of Neurology, Qilu Hospital of Shandong University. His research interests include small cerebrovascular disease, cognitive impairment, encephalitis and anxiety, and depression.



JIANPING QIAO received the B.S. degree in electronic engineering from the University of Petroleum, China, in 2003, and the Ph.D. degree in communication and information system from Shandong University, Jinan, China, in 2008.

She is currently an Associate Professor with the School of Physics and Electronics, Shandong Normal University, Jinan. Her current research interests include MRI image processing and analysis, brain structure and function analysis, and brain network analysis.

• • •