Automatic Diagnosis of Alzheimer’s Disease and Mild Cognitive Impairment Based on CNN+SVM Networks with End-to-end Training

Minglang Sun, Zhe Huang and Chengan Guo*
School of Information and Communication Engineering
Dalian University of Technology, Dalian, 116023, China
Emails: suminglang@mail.dlut.edu.cn, hzdlmu@mail.dlut.edu.cn and cguo@dlut.edu.cn

Abstract—Alzheimer’s disease (AD) is an irreversible neurodegenerative disease and, at present, once it has been diagnosed, there is no effective curative treatment. Accurate and early diagnosis of Alzheimer’s disease is crucial for improving the condition of patients since effective preventive measures can be taken in advance to delay the onset time of the disease. Fluorodeoxyglucose positron emission tomography (FDG-PET) is an effective biomarker of the symptom of AD’s, and has been used as medical imaging data for diagnosing AD’s. Mild cognitive impairment (MCI) is regarded as an early symptom of AD’s, and it has been shown that MCI also has a certain biomedical correlation with FDG-PET. In this paper, we explore how to use 3D FDG-PET images to realize the effective recognition of MCI’s, and thus achieve the early prediction of AD’s. This problem is then taken as the classification of three categories of 3D FDG-PET images, including MCI, AD and NC (normal controls). In order to get better classification performance, a novel network model is proposed in the paper based on 3D convolution neural networks (CNN) and support vector machines (SVM) by utilizing both the excellent abilities of CNN in feature extraction and SVM in classification. In order to make full use of the optimal property of SVM in solving binary classification problems, the three-category classification problem is divided into three binary classifications, each binary classification being realized with a CNN+SVM network. Then the outputs of the three CNN+SVM networks are fused into a final three-category classification result. An end-to-end learning algorithm is developed to train the CNN+SVM networks and a decision fusion strategy is exploited to realize the fusion of the outputs of three CNN+SVM networks. Experimental results obtained in the work with comparative analyses confirm the effectiveness of the proposed method.

Keywords—Alzheimer’s disease, mild cognitive impairment, FDG-PET, CNN+SVM, end-to-end training, decision fusion

I. INTRODUCTION

Alzheimer’s disease (AD) is a chronic neurodegenerative disease affecting the elderly people. Patients with AD generally result in obvious cognitive dysfunction and memory decline. As the disease usually causes cell necrosis and tissue loss, it is regarded as the 6th leading cause of death in the U.S. [1,2]. The number of AD patients in the world will reach 300 million by 2050 [3]. Unfortunately, since there is no effective curative treatment to reverse AD at present, early diagnosis and treatment for AD’s is vital for patient care and development of future treatment. Mild cognitive impairment (MCI) is often considered as a clinical prodromal phase of AD, usually occur before the AD dementia that has minimal impact on daily life [4]. Therefore, the MCI detection plays an important role in early detection of AD. As the metabolic rate and structure of the brain changes accordingly with the progression of AD, multimodality neuroimages, e.g. positron emission tomography (PET) and magnetic resonance images (MRI) can be utilized to quantify the changes and further applied for computer-aided diagnosis (CAD) [5,6].

For AD detection, many studies based on PET images have been presented. Gray et al [7] segmented the PET image into 83 anatomical regions of interests (ROI), and employed a support vector machine (SVM) trained by the mean signal intensity in these regions for classification. Garali et al [8] extracted the entropy of histograms of anatomical ROI as features, and utilized both SVM and random forest to classify the features to get the final decision. Silveira and Marques [9] proposed a boosting classification framework for voxel-wise features, which integrated several classifiers to diagnose AD. Cabral et al [10] used an ensemble of SVM and random forest to perform classification on different feature subsets.

In recent years, with the rapid development of deep learning and the excellent performance on image classification, amounts of studies based on deep learning have been proposed for AD detection. Different from the conventional methods which need to extract features manually, the methods based on deep learning can automatically capture latent features from images without aiming at specific modality. Wang et al [11] proposed an eight-layer convolutional neural network (CNN) with leaky rectified linear unit and maxpooling for AD classification, in which 2D slice of 3D MRI is employed as the inputs of CNN. Ding et al [12] introduced the inception v3 that stacks 11 inception modules [13] into the method for AD classification with the 4 × 4 grid images generated from the 3D PET as inputs.

Although the above mentioned methods show the effectiveness in AD classification, one of the shortcomings of the methods with 2D slice inputs is that the information in all dimensions of the 3D image is not fully utilized. In order to solve this problem, Liu et al [14] combined the CNN and recurrent neural network (RNN) for AD classification, in which 2D CNN is used to capture the intra-slice features, and RNN is used to learn inter-slice features. The features of image slices produced by 2D CNNs were fed into the input of each gate recurrent unit, and the final results were obtained by fusing the prediction scores from three direction of 3D FDG-PET. Huang et al [15] constructed a 3D VGG variant model based on single modality for AD diagnosis and achieved multi-modality detection by concatenating the multi-modality features obtained from MRI and PET images. In addition, the experimental results
of [15] showed that hippocampus segmentation is not necessary for improving the performance of CNN-based classification method.

Besides, most of the existing studies on the topic aim at solving the binary classification problems, such as AD vs. NC or MCI vs. NC. However, in order to achieve the early prediction of AD symptoms, an MCI sample must be distinguished from AD or NC samples as much as possible, because MCI is a transition state from NC to AD, and it is more difficult to be correctly identified compared with the identification of AD and NC. One way to solve this problem is to directly build a 3-category classifier, but this is usually not able to achieve the excellent enough performance, thus more attention needs to be paid on the identification of MCI’s.

In view of the optimal property of SVM in solving binary classifications and the powerful feature extraction ability of deep CNNs, in this paper, we integrate the advantages of the two methods and design a hybrid classification network based on integration of CNN and SVM models. Moreover, an end-to-end training algorithm is developed for further fine-tuning the hybrid system, and a decision fusion strategy is proposed to perform the fusion of the outputs of three CNN+SVM networks. Extensive experiments have been conducted in the work and the experimental results show that the proposed approach achieves outstanding performance, compared with other state of the art methods.

The sequel of this paper is organized as follows: Section II describes the related work, Section III presents the detailed description of the proposed method and the databases used in the work, and Section IV gives the experimental results and performance analysis. Finally, Section V draws conclusions of the contributions made in the paper.

II. RELATED WORK

A. Techniques of Boosting the Performance for Diagnosis of AD Based on CNN

In [16], Liu et al. proposed a CNN-based model for AD automatic diagnosis and introduced some techniques on designing the CNN model. Generally, in order to ease the training of model, batch normalization (BN) technique is widely applied in most CNN models. However, BN has poor performance under small batch size, which usually occurs in CNN models. Consequently, instead of using BN, instance normalization (IN) is used to replace BN in [16], which maintains excellent performance on small batch size [17]. Different from most mainstream framework [18,19] for image classification, small kernel size and stride were used in the first layer and obtained significant improvement for AD classification. In their architecture, increasing the width of the network achieved better performance than the depth. Besides that, since the brain shrinks with age and Alzheimer's disease may be positively related to it, age information may contribute to AD diagnose. In order to integrate age information, age value may be positively related to it, age information may contribute to AD diagnose. In order to integrate age information, age value was encoded into a vector and combined with the output of the convolutional layers in the paper.

B. Support Vector Machine

SVM is proposed originally by Vapnik and its effectiveness has been proved in binary classification problems [20] and purpose of SVM is to find a separation hyperplane, which maximize the distances between the margins of two kinds of classifications. The main algorithm of SVM model are as follows [20].

For n-sample label pairs \((x_i, y_i)\), \(x_i \in \mathbb{R}^{d} \), \(i = 1, ..., n\) \(y_i \in \{-1, 1\}\), the objective function of SVM is defined by:

\[
L(w,b,\alpha,\xi) = \frac{1}{2} \|w\|^2 + C \sum_i \xi_i - \sum_i \{\alpha_i[y_i(w^T x_i + b) - 1]\}
\]  

where \(i = 1, ..., n\), \(w \in \mathbb{R}^d\) is the coefficient vector, \(b\) is the bias term, \(\alpha \geq 0\) is Lagrange multiplier, \(\xi\) is a relaxation parameter which allows some misclassification, \(C \geq 0\) is a penalty parameter used to control the degree of penalty for misclassification.

In order to optimize the SVM by minimizing its objective function, (1) can be solved by the following dual problem:

\[
Q(\alpha) = \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j K(x_i, x_j)
\]

Subjected to: \(0 \leq \alpha_i \leq C\), \(\sum_i \alpha_i y_i = 0\)

where, \(i, j = 1, ..., n\), and \(K(x_i, x_j)\) is the kernel function.

Kernel function, \(K(x_i, x_j)\), is introduced to achieve linear classification after mapping nonlinear input vectors onto higher dimensional feature space without increasing computational complexity. Some common kernel functions are given below:

Polynomial kernel function:

\[
K(x_i, x_j) = [(x_i \cdot x_j) + 1]^q
\]

where \(x\) is the input vector, \(x_i\) denotes the support vector of SVM, \(q\) is the order of polynomial.

Gaussian Radial Basis Function:

\[
K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}
\]

where \(\sigma\) is the standard deviation.

It can be seen that by transforming the original problem into a dual problem, the computational complexity no longer depends on the input dimension, but on the number of samples, especially on the number of support vectors in the samples. These characteristics make it possible to deal with high dimensional problems effectively. After solving the dual problem, all the parameters of \(Q\) are obtained and the bias and separating hyperplane can be obtained by:

\[
b = \frac{1}{\|S\|} \sum_i \{y_i - \sum_j \alpha_j y_j K(x_i, x)\} \]

280
where \( \| S \| \) indicates the number of non-zero parameters of \( \alpha \)'s, and \( \text{sgn}(\cdot) \) denotes the sign function.

### III. PROPOSED METHOD

#### A. Overall Scheme of the Proposed Method

In view of the excellent feature extraction ability of CNN and the optimal property of SVM for binary classifications, a hybrid model integrated with CNN and SVM networks is designed in the paper. The structure of the proposed hybrid model is shown in Fig.1 which is composed of two modules, a feature extraction module based on a 3 dimensional CNN (called 3DCNN) and a SVM-based classification module. Inspired by [16], the 3DCNN model is redesigned here in accordance to the purpose of this paper and it is adopted as a baseline to extract deep features from the input 3D FDG-PET images, and the SVM model is exploited to integrate with the 3DCNN to get better classification performance by utilizing the features extracted from the 3DCNN. In order to optimize the 3DCNN+SVM network, an end-to-end training strategy is designed to train the hybrid learning model. As the optimal classification performance can be achieved by SVM for binary classifications, to improve the performance of 3-category classification problem, we divide the 3-category classification problem into three binary classification problems. The overall structure of this three-category classification system is shown in Fig.2, in which it consists of three 3DCNN+SVM networks, each binary classification being realized with one 3DCNN+SVM network. In order to obtain a final classification result as good as possible, a decision fusion strategy is proposed to fuse the outputs of three 3DCNN+SVM classifiers. The details of the proposed classification system will be given in the sequel subsections.

#### B. 3DCNN-based Feature Extraction Module

CNN is widely used in the field of computer vision currently [21]. Different from conventional methods that extract features manually, CNN can automatically learn features through an end-to-end training process. In order to extract 3D image features effectively, a 3DCNN network is designed in the work and its structure is described in TABLE I, which consists of convolution layer, maxpooling layer, channel attention layer, global average pooling layer, and softmax layer. Different from commonly used classification models, such as 3D ResNet and 3D DenseNet, dimension reduction is not performed after the first layer of the model so as to learn better lesion feature from the 3D FDG-PET images. Due to poor performance of batch normalization in small batch size, instance normalization which maintains excellent performance on small batch size is used in our work. In addition, the channel attention in CBAM is applied in our model to focus on the key feature maps [22]. The generated attention vector \( W \) from channel attention can be described as:

\[
W(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F)))
\]

where \( F \) denotes the input feature map, \( \sigma(\cdot) \) is the sigmoid function, and MLP indicates the multilayer perceptron.

In addition, the dropout and label smoothing technology are employed to alleviate overfitting [23, 24]. The outputs of the feature extraction module obtained after global average pooling (GAP) layer are used for subsequent classification.

#### TABLE I. THE ARCHITECTURE OF 3DCNN DESIGNED IN THE PAPER

| Layer ID | Layer    | Kernel number | Kernel size/stride | Output size |
|----------|----------|---------------|--------------------|-------------|
| 0        | Input    |               |                    | 1 × 80 × 100 × 76 |
| 1        | Conv1    | 32            | (1, 1, 1)/1        | 32 × 80 × 100 × 76 |
| 2        | Conv2    | 64            | (3, 3, 3)/1        | 64 × 80 × 100 × 76 |
| 3        | MaxPool3D| (2, 2, 2)/2    | 64 × 40 × 50 × 38 |
| 4        | Conv3    | 128           | (3, 3, 3)/1        | 128 × 40 × 50 × 38 |
| 5        | Attention|               | 128 × 40 × 50 × 38 |
| 6        | Maxpool3D| (2, 2, 2)/2    | 128 × 20 × 25 × 19 |
| 7        | Conv4    | 256           | (3, 3, 3)/1        | 256 × 20 × 25 × 19 |
| 8        | Attention|               | 256 × 20 × 25 × 19 |
| 9        | Maxpool3D| (2, 2, 2)/2    | 256 × 10 × 12 × 9  |
| 10       | Conv5    | 512           | (3, 3, 3)/1        | 512 × 10 × 12 × 9  |
| 11       | Attention|               | 512 × 10 × 12 × 9  |
| 12       | Maxpool3D|               | 512 × 5 × 6 × 4    |
| 13       | Conv6    | 512           | (3 × 3 × 3)/1      | 512 × 3 × 4 × 2    |
| 14       | GAP      | 512           | 512 × 1 × 1 × 1    |
| 15       | Flatten  | 512           |                    |
| 16       | Softmax  | 2             |                    |

#### C. SVM-based Classification Module with an End-to-end Training Strategy

Due to the excellent performance of SVM in binary classification tasks, it is utilized in the work for solving our problem. The feature vector after global average pooling of CNN is regarded as the input of SVM. It is known that the performance of SVM depends on the support vectors. Once the CNN is trained, the support vectors are fixed. Under these
support vectors, the trained SVM can get the optimal solution for the data with the same statistical structure. However, the feature vectors from CNN may not be the most suitable inputs for the SVM at this stage. Basing on the assumption, we present an end-to-end training strategy for the 3DCNN+SVM system to get better solution.

Since the SVM is trained with all the inputs, we need to train the CNN and SVM in advance. Firstly, we train the 3DCNN network described above with 3D FDG-PET images to obtain CNN feature vectors. Secondly, we train a SVM model with Gaussian kernel function using the CNN features as input vectors. Finally, we connect both the trained CNN and SVM, and fine-tune the 3DCNN+SVM system by an end-to-end training strategy.

As shown in (6), a non-derivable sign function is employed to binarize the values of the linear output of SVM to obtain finally prediction, which can be described as:

\[ y = \text{sgn}(s) \]  

where \( y \) is the output of SVM, \( s \) is the linear output of SVM.

Since the BP algorithm cannot be performed by using non-differentiable sign function, the sign function is replaced with a differentiable softmax function and the linear output of SVM is taken as the input of the softmax function.

Due to that the output of sign function is 1 or -1, the influence of the linear output value of SVM is ignored. In general, higher value of the output in the classification indicates higher confidence that the input belongs to the positive samples. In order to correlate the confidence with the inf luence of the linear output value of SVM is ignored. In general, higher value of the output in the classification indicates higher confidence that the input belongs to the positive samples. In order to correlate the confidence with the output value, we use softmax function to quantify the output. Since SVM only has one output, we use the linear value \( s \) together with its opposite value, -\( s \), as the inputs of the softmax layer. In this way, the probability that an input belongs to the positive set and negative set can be computed as follows, respectively:

\[ q(x \in d_p) = \frac{e^s}{e^s + e^{-s}} \]  

\[ q(x \in d_n) = \frac{e^{-s}}{e^s + e^{-s}} \]

where \( x \) is the input image, \( s \) is the linear output value of SVM, \( d_p \) and \( d_n \) denotes that \( x \) belongs to the positive class in probability \( q(x \in d_p) \), and negative class in probability \( q(x \in d_n) \), respectively.

Therefore, we use the cross-entropy loss to optimize the proposed model. The cross-entropy loss is defined as:

\[ H(p,q) = - \sum_{i=1}^{n} p(x_i \in d_p) \log q(x_i \in d_p) + (1 - p(x_i \in d_p)) \log (1 - q(x_i \in d_p)) \]

where \( p \) is the label function that defines as \( p = 1 \) if \( x \) is positive else \( p = 0 \), \( n \) indicates the total numbers of the training samples.

With the loss function increasing in training procedure, \( s \) tends to increasing for positive samples and -\( s \) also tends to increasing for negative samples. Therefore, this loss function can not only realize the BP algorithm, but also enhance the distance between the positive and negative samples.

In view of the fact that, when a SVM has been got trained, its outputs are the optimal under the same inputs, we reserve some training samples as verification samples in the pre-training stage, and add these verification samples in the fine-tuning stage. Meanwhile, the SVM is fixed to prevent it from degenerating to a fully connected network on the stage of training CNN. In the fine-tuning stage, CNN and SVM are trained alternately when training CNN, the weights of SVM are locked without updated, while training SVM, the weights of CNN are locked without updated. Details for these operation steps are as follows:

(i) Initialize a 3DCNN and a SVM to be trained, and divide the PET dataset into 3 subsets (training set, verification set, and test set).
(ii) Train the 3DCNN by using the samples in training set until converged, then use the converged 3DCNN to extract the feature vector output from its last pooling layer using all the samples in the training set and in the verification set as input.
(iii) Train the SVM by using the extracted feature vectors as training samples obtained by using the input samples in the training set in step (ii), until the SVM converged.
(iv) Construct a 3DCNN+SVM network using the trained 3DCNN and SVM, and replace sign function with softmax function as described in (9) and (10).
(v) Fine-tune the 3DCNN+SVM network by using the samples both in the training set and in the verification set and the loss function computed according to (11), with the weights of the SVM fixed (without updated), until the 3DCNN converged basically.
(vi) Re-train the SVM by using the extracted feature vectors output from the 3DCNN obtained in step (v) without updating the 3DCNN, until the SVM converged basically.
(vii) Repeat the steps (iv)-(vi), until the whole 3DCNN+SVM network converged.
(viii) Test the trained 3DCNN+SVM network by using the samples in the test set.

Three 3DCNNi+SVM, networks (\( i=1,2,3 \)) can be built up, as shown in Fig.2, by using the above training algorithm for conducting three binary classification tasks of AD vs. NC, NC vs. MCI, and MCI vs. AD, with the corresponding training samples.

D. Decision Fusion Strategy of Three Classification Results

At present, most of the existing studies related to AD are devoted to binary classification tasks, such as AD vs. NC and MCI vs. NC. However, in practical applications, a robust 3-category classification model is crucial. In general, this problem can be solved directly by designing a 3-category classifier, but this is often not able to achieve an optimal performance. In order to obtain better result, we divide the 3-category classification problem into three binary classification problems, and propose a decision fusion strategy to realize three-category classification.
Before making a final decision fusion, three 3DCNN+SVM networks need to be trained in advance for performing the binary classifications of AD vs. NC, MCI vs. NC, and AD vs. MCI, in which each 3DCNN+SVM model is constructed by the proposed method described above. Suppose that the three CNN+SVM-based binary classifiers have been trained, the decision fusion strategy will be operated in the following steps:

(i) For a 3D FDG-PET image to be classified, it is fed into the three 3DCNN+SVM networks \( (i=1,2,3) \) respectively, and three classification results can be obtained;

(ii) If the results of two classification models belong to the same category, the category is regarded as the final classification result;

(iii) If all the three classification results are different, the final decision is made according to the absolute value, \( |s_i| \), of the linear output of the SVM\( _i \) (\( i=1,2,3 \)) as follows:

\[
k = \arg \max (|s_i|)\tag{12}
\]

Then final classification result is selected as the binary classification result of the \( k \)-th 3DCNN+SVM network (i.e., the output of the SVM\( _k \)).

IV. EXPERIMENTS

A. Database and Implementation Settings

Data used in the work were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database [25] launched in 2003 and ADNI has been committed to tracking the progression of AD through biomarkers and clinical assessments. Identifying sensitive and specific markers of early AD progression in the database can help researchers and clinicians develop new treatments, monitor the effectiveness and reduce the cost of clinical trials.

In this work, we adopt 2706 3D FDG-PET images from 959 ADNI participants, including 267 AD subjects, 340 MCI subjects, 352 NC subjects. TABLE II presents the demographic details of the studied subjects in the work, where MMSE is the abbreviation of the Mini-mental State Examination. The voxels outside the brain are removed from the FDG-PET images, and the images are cropped to size of \( 80 \times 100 \times 76 \) for classification.

TABLE II. DEMOGRAPHIC CHARACTERISTICS OF THE STUDIED SUBJECTS

| Diagnosis | Number | Age | Gender (F/M) | MMSE |
|-----------|--------|-----|-------------|------|
| AD        | 514    | 75.98±7.62 | 305/209 | 19.26±5.64 |
| MCI       | 1247   | 76.47±7.54 | 809/438 | 22.83±6.56 |
| NC        | 945    | 76.99±5.95 | 544/405 | 27.83±3.63 |

All the models and algorithms adopted in the work have been implemented and all the experiments are conducted by using Python on a CPU+GPU platform with the CPU of Intel®Core™ i7-7700@3.60 GHz and the GPU of NVIDIA GeForce GTX 1080Ti. Five-fold cross-validation is performed in the experiments, where the data set is divided into 5 equal parts in which 1 part is used as the testing data and 4 parts are used as training data with 1 part of them as verification data. And the experiments are conducted 5 times in turn, and the mean values of the results of 5 trials are used as final indexes of the method. The data are strictly divided according to patient’s IDs to ensure that the image samples of the same person will not be put into different data sets, i.e., the FDG-PET images of one participant are put into only one part in the data partition. The stochastic gradient descent (SGD) algorithm is utilized to minimize the loss function in training the proposed model. The batch size is set to 4 3D-images and the weights of the network are updated every four batches for better convergence in the training process.

In order to evaluate the performance of the proposed method, we use 4 technical indexes for evaluation, including accuracy (ACC), sensitivity (SEN), specificity (SPE), and AUC (area under ROC curve). The ACC, SEN, and SPE are the proportion of correct predictions among all samples, positive samples, and negative samples respectively. The AUC is obtained by computing the area under receiver operating characteristic curve (ROC) which is the curve to describe the relationship between the true positive rate (TPR) and the false positive rate (FPR) under varied threshold settings.

B. Evaluation of the Proposed Method Applied to Binary Classification

In this section, experiments are conducted for the proposed 3DCNN+SVM classification method, and also for the other state-of-the-art methods proposed in [12], [14], and [15] for comparisons, respectively. The methods proposed in the cited literatures were originally for solving a binary classification problem, such as for classifying the samples of AD vs. NC, or MCI vs. NC. For our proposed method, since a single 3DCNN+SVM model with end-to-end training is also proposed for solving a binary classification problem, we just need to use a single 3DCNN+SVM network to perform the classification without needing three such networks. As for the experimental results given in the cited literatures, they were obtained by using different data partition under different experiment settings, in order to make a fair comparison, these algorithms are re-implemented by using the same FDG-PET data under the same experiment settings as in ours in the paper.

TABLE III. EVALUATION OF THE PROPOSED 3DCNN+SVM WITH E2E APPLIED TO BINARY CLASSIFICATION OF AD VS. NC SAMPLES (%)

| Method          | ACC  | SEN  | SPE  | AUC  |
|-----------------|------|------|------|------|
| Ding et al. [12]| 86.1 | 86.9 | 85.6 | 90.4 |
| Liu et al. [14] | 89.2 | 89.7 | 88.9 | 92.1 |
| Huang et al. [15]| 88.6 | 89.0 | 88.5 | 91.8 |
| 3DCNN+SVM with E2E | 90.1 | 90.8 | 89.7 | 93.4 |

TABLE IV. EVALUATION OF THE PROPOSED 3DCNN+SVM WITH E2E APPLIED TO BINARY CLASSIFICATION OF MCI VS. NC SAMPLES (%)

| Method          | ACC  | SEN  | SPE  | AUC  |
|-----------------|------|------|------|------|
| Ding et al. [12]| 72.8 | 73.8 | 71.4 | 79.3 |
| Liu et al. [14] | 74.2 | 75.2 | 72.9 | 80.1 |
| Huang et al. [15]| 73.5 | 74.6 | 72.1 | 79.8 |
| 3DCNN+SVM with E2E | 75.9 | 77.3 | 74.1 | 80.7 |
training proposed in the paper. From the results shown in the tables, one can see that the proposed method performs the best among the other ones, and its effectiveness is confirmed by the experiments.

C. Evaluation of the Proposed Method Applied to 3-Category Classification

As mentioned before, in order to solve the early prediction of AD symptoms, we design a hybrid 3-category classification system by integrating three binary 3DCNN+SVM classifiers with an optimal decision fusion scheme. In this subsection, we present the experimental results to evaluate this 3-category classification system by using the 3D FDG-PET images from MCI, AD and NC subjects. In order to demonstrate the effectiveness of the proposed method, a CNN-based state-of-the-art method proposed by Liu et al [16] is implemented in the paper for comparison. The CNN-based method [16] was also for solving the same 3-category classification problem as in this paper but using the MRI data originally. In this work we re-implement the CNN-based model proposed in [16], train and test by using the same 3D FDG-PET images as used in the paper. Table 5 shows the experiment results, in which the experimental results of 3D-CNN are also included that are obtained by using a three dimensional CNN network with the same structure as the 3DCNN of our 3DCNN+SVM model but with the SVM substituted by an output layer of three softmax nodes. This 3D-CNN model is also a 3-category classifier, it is trained and tested by using the same data as the other models, and also used for performance comparison in the experiment.

From the results shown in Table V, one can see that the proposed hybrid 3-category classification system obtains a significant improvement on all its classification performance, compared with the directly trained 3-category CNN-based classification method proposed in [16]. Compared with 3D-CNN method, the proposed classification method also gets significant improvements on the overall performance, except for a slight degradation on the MCI category. This shows that the proposed hybrid classification approach outperforms the directly trained CNN-based classification method.

C. Evaluation of the Proposed Method Applied to 3-Category Classification

As mentioned before, in order to solve the early prediction of AD symptoms, we design a hybrid 3-category classification system by integrating three binary 3DCNN+SVM classifiers with an optimal decision fusion scheme. In this subsection, we present the experimental results to evaluate this 3-category classification system by using the 3D FDG-PET images from MCI, AD and NC subjects. In order to demonstrate the effectiveness of the proposed method, a CNN-based state-of-the-art method proposed by Liu et al. [16] is implemented in the paper for comparison. The CNN-based method [16] was also for solving the same 3-category classification problem as in this paper but using the MRI data originally. In this work we re-implement the CNN-based model proposed in [16], train and test by using the same 3D FDG-PET images as used in the paper. Table 5 shows the experiment results, in which the experimental results of 3D-CNN are also included that are obtained by using a three dimensional CNN network with the same structure as the 3DCNN of our 3DCNN+SVM model but with the SVM substituted by an output layer of three softmax nodes. This 3D-CNN model is also a 3-category classifier, it is trained and tested by using the same data as the other models, and also used for performance comparison in the experiment.

Table V: Evaluation of the Proposed Method Applied to 3-Category Classification in Terms of ACC (%)

| Method          | AD   | MCI  | NC   | Average |
|-----------------|------|------|------|---------|
| Liu et al. [16] | 66.5 | 67.1 | 66.1 | 66.6    |
| 3D-CNN          | 66.1 | 67.7 | 64.7 | 65.9    |
| Proposed        | 71.7 | 67.3 | 71.3 | 69.5    |

D. Ablation Experiments

In this section, we conduct ablation experiments to evaluate the effectiveness of the proposed 3DCNN+SVM classification method, by using the 3DCNN network with an output layer of 2 softmax nodes that is also the baseline of the ablation experiments, the direct combination model of a 3DCNN and an SVM (i.e., 3DCNN+SVM) that the two models are trained separately without end-to-end fine tuning jointly, and the model of 3DCNN+SVM plus an end-to-end fine tuning process.

Table VI: Ablation Studies of the Proposed 3DCNN+SVM Model Applied to Binary Classification of AD vs. NC (%)

| Method                  | ACC  | SEN  | SPE  | AUC   |
|-------------------------|------|------|------|-------|
| 3DCNN                   | 89.3 | 89.9 | 89.0 | 91.9  |
| 3DCNN+SVM               | 89.7 | 90.6 | 89.2 | 92.6  |
| 3DCNN+SVM+E2E           | 90.1 | 90.8 | 89.7 | 93.4  |

Table VII: Ablation Studies of the Proposed 3DCNN+SVM Model Applied to Binary Classification of MCI vs. NC (%)

| Method                  | ACC  | SEN  | SPE  | AUC   |
|-------------------------|------|------|------|-------|
| 3DCNN                   | 74.9 | 76.3 | 73.0 | 78.3  |
| 3DCNN+SVM               | 75.4 | 76.8 | 73.5 | 80.1  |
| 3DCNN+SVM+E2E           | 75.9 | 78.3 | 74.1 | 80.7  |

In addition, Fig. 3 and Fig. 4 display the comparison of the ROC curves on AD vs. NC and MCI vs. NC for above ablation experiments.
In this paper, we proposed a new classification system for early automatic diagnosis of AD symptoms based on 3DCNN and SVM, in which, the original 3-category classification problem is divided into three binary classification problems, each binary classification is realized with a 3DCNN+SVM model. Furthermore, an end-to-end learning algorithm is developed for training the 3DCNN+SVM networks, and an optimal decision fusion scheme is proposed to fuse the outputs of three 3DCNN+SVM classifiers based on the criteria of majority voting. By using these methods, the advantages of both CNN and SVM models can be fully utilized, thus the overall performance of the system can be significantly improved. Experimental results obtained in the paper confirm the effectiveness of the proposed approach that outperforms the existing start-of-the-art methods in terms of the class accuracy, sensitivity, specificity, and area under ROC.

It is noticed that, from the experimental results obtained in the paper, the classification performance of MCI samples still leaves some room for further improvement, and the correct identification of this category samples is crucial for the early diagnosis of AD. Therefore, more effective method is needed to be developed to overcome this shortage, which will be the future research direction of the paper.

REFERENCES

[1] G. McKhann, D. Drachman, M. Folstein, et al, “Clinical diagnosis of Alzheimer’s disease: Report of the Nincds-Adrd work group under the auspices of department of health and human services task force on Alzheimer’s disease,” Neurology, vol. 34, no.7, pp.939-939, 1984.

[2] National Center for Health Statistics. Deaths and mortality, May 2017. URL http s://www.cdc.gov/nchs/fastats/deaths.html. [Online; posted 07-May-2017].

[3] X. F. Zhu, H. Suk, Y. H. Zhu, et al, “Multi-view Classification for Identification of Alzheimer’s Disease,” in MLMI, pp. 255–262, 2015.

[4] R. C. Petersen, G. E. Smith, S. C. Waring, et al, “Mild cognitive impairment: clinical characterization and outcome,” Archives of neurology, vol. 56, no. 3, pp. 303–308, 1999.

[5] H. Matsuda, “The role of neuroimaging in mild cognitive impairment,” Neuroapathology, vol. 27, no. 6, pp. 570–577, 2010.

[6] L. Mosconi, W. H. Tsui, K. Herholz, et al, “Multicenter standardized 18F-FDG PET diagnosis of mild cognitive impairment, Alzheimers’ disease, and other dementias,” J. Nucl. Med., vol. 49, no. 3, pp. 390–398, 2008.

[7] K. R. Gray, R. Wolz, S. Kelhaninejad, et al, “Regional analysis of FDG-PET for use in the classification of Alzheimer’s disease,” in ISBI, Chicago, IL. Mar.30-Apr.02, 2011, pp. 1082-1085.

[8] I. Garali, M. Adel, S. Bourremane, and E. Guedj, “Region-based brain selection and classification on PET images for Alzheimer’s disease computer aided diagnosis,” in ICIP, 2015, pp. 1473-1477.

[9] M. Silveira and J. S. Marques, “Boosting Alzheimer’s disease diagnosis using PET images,” in ICPR 2010, pp. 23-26.

[10] C. Cabral and M. Silveira, “Classification of Alzheimer’s disease from FDG-PET images using favorite class ensembles,” in EMBC, Osaka, Japan, Jul. 03-07, 2013, pp. 2477–2480.

[11] S. H. Wang, P. Phillips, Y. Sui, et al, “Classification of Alzheimer’s disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling,” J. Nucl. Med., vol. 42, no. 5, p. 85, 2018.

[12] Y. M. Ding, J. H. Sohn, M. G. Kawczynski, et al, “A deep learning model to predict a diagnosis of Alzheimer disease by using 18F-FDG PET of the brain.” Radiology, vol. 290, no. 2, pp. 456-464, 2018.

[13] C. Szegedy, V. Vanhoucke, S. Ioffe, et al, “Rethinking the inception architecture for computer vision,” in CVPR, 2016, pp. 2818–2826.

[14] M. H. Liu, D. N. Cheng, and W. W. Yuan, “Classification of Alzheimer’s disease by combination of convolutional and recurrent neural networks using FDG-PET images,” Frontiers in Neuroinformatics, vol. 12, 2018, DOI. 10.3389/fninf.2018.00105.

[15] Y. Huang, J. Xu, Y. Zhou, et al, “Diagnosis of Alzheimer’s disease via multi-modality 3d convolutional neural network,” Frontiers in Neuroscience, vol.13, 2019, DOI. 10.3389/fnins.2019.00509.

[16] S. Liu, C. Yadav, C. Fernandez-Granda, et al, “On the design of convolutional neural networks for automatic detection of Alzheimer’s disease,” in PMLR, 2020, pp. 184-201.

[17] D. Ulyanov, A. Vedaldi, and V. Lempitsky, “Instance normalization: The missing ingredient for fast stylization,” bioRxiv, 2016. [Online]. Available: https://arxiv.org/abs/1607.08022.

[18] G. Huang, Z. Liu, V.D.M. Laurens, et al, “Densely connected convolutional networks,” in CVPR, 2016, pp. 4700-4708

[19] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in CVPR, 2016, pp. 770-778.

[20] C. Cortes and V. Vapnik, “Support vector network,” Machine learning, vol. 20, no. 3, pp. 273-297, 1995.

[21] Y. Lecun and L. Bottou, “Gradient-based learning applied to document recognition,” in Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[22] S. Woo, J. Park, J.Y. Lee, and I. S. Kweon, “Cbam: Convolutional block attention module,” in ECCV, 2018, pp. 770-778.

[23] N. Srivastava, G. Hinton, A. Krizhevsky, et al, “Dropout: A simple way to prevent neural networks from overfitting,” in J. Mach. Learn. Res, vol. 15, no. 1, pp. 1929-1958, 2014.

[24] R. Muler, S. Kornblith, and G. Hinton, “When does label smoothing help?” bioRxiv, 2019. [Online]. Available: https://arxiv.org/abs/1906.02629.

[25] ADNI database: http://adni.loni.usc.edu/.

Fig. 4. ROC curves of the ablation experiments on MCI vs. NC

V. CONCLUSIONS

In this paper, we proposed a new classification system for early automatic diagnosis of AD symptoms based on 3DCNN and SVM, in which, the original 3-category classification problem is divided into three binary classification problems, each binary classification is realized with a 3DCNN+SVM model. Furthermore, an end-to-end learning algorithm is developed for training the 3DCNN+SVM networks, and an optimal decision fusion scheme is proposed to fuse the outputs of three 3DCNN+SVM classifiers based on the criteria of majority voting. By using these methods, the advantages of both CNN and SVM models can be fully utilized, thus the overall performance of the system can be significantly improved. Experimental results obtained in the paper confirm the effectiveness of the proposed approach that outperforms the existing start-of-the-art methods in terms of the class accuracy, sensitivity, specificity, and area under ROC.

It is noticed that, from the experimental results obtained in the paper, the classification performance of MCI samples still leaves some room for further improvement, and the correct identification of this category samples is crucial for the early diagnosis of AD. Therefore, more effective method is needed to be developed to overcome this shortage, which will be the future research direction of the paper.