Computers in Biology and Medicine 134 (2021) 104478

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Multiclass diagnosis of stages of Alzheimer's disease using linear discriminant analysis scoring for multimodal data☆

Weiming Lin [a,b], Qinquan Gao [c,d], Min Du [c,e], Weisheng Chen [f], Tong Tong [c,g,*]

[a] School of Opto-Electronic and Communication Engineering, Xiamen University of Technology, Xiamen, 361024, China
[b] Fujian Key Laboratory of Communication Network and Information Processing, Xiamen University of Technology, Xiamen, 361024, China
[c] College of Physics and Information Engineering, Fuzhou University, Fuzhou, 350116, China
[d] Imperial Vision Technology, Fuzhou, 350001, China
[e] Fujian Provincial Key Laboratory of Eco-industrial Green Technology, Wuyi University, Wuyishan, 354300, China
[f] Department of Thoracic Surgery, Fujian Cancer Hospital, Fuzhou, 350001, China
[g] Fujian Key Lab of Medical Instrumentation & Pharmaceutical Technology, Fuzhou University, Fuzhou, 350116, China

ARTICLE INFO

Keywords:
Alzheimer's disease
Mild cognitive impairment
Multiclass
Multimodal
Linear discriminant analysis
Extreme learning machine

ABSTRACT

Alzheimer's disease (AD) is a progressive neurodegenerative disease, and mild cognitive impairment (MCI) is a transitional stage between normal control (NC) and AD. A multiclass classification of AD is a difficult task because there are multiple similarities between neighboring groups. The performance of classification can be improved by using multimodal data, but the improvement could be limited with inefficient fusion of multimodal data. This study aims to develop a framework for AD multiclass diagnosis with a linear discriminant analysis (LDA) scoring method to fuse multimodal data more efficiently. Magnetic resonance imaging, positron emission tomography, cerebrospinal fluid biomarkers, and genetic features were first preprocessed by performing age correction, feature selection, and feature reduction. Then, they were individually scored using LDA, and the scores that represent the AD pathological progress in different modalities were obtained. Finally, an extreme learning machine-based decision tree was established to perform multiclass diagnosis using these scores. The experiments were conducted on the AD Neuroimaging Initiative dataset, and accuracies of 66.7% and 57.3% and F1-scores of 64.9% and 55.7% were achieved in three- and four-way classifications, respectively. The results also showed that the proposed framework achieved a better performance than the method that did not score multimodal data and the methods in previous studies, thereby indicating that the LDA scoring strategy is an efficient way for multimodalities fusion in AD multiclass classification.

1. Introduction

As people aged above 65 years are at a high risk of developing Alzheimer's disease (AD) [1], AD has become one of the major concerns related to the health of elderly people. Patients with AD often suffer from memory loss, language disorders, and disorientation. Consequently, a lot of time and energy is required to take care of patients with AD, which results in considerable stress on the families or caregivers. There are still no efficient drugs or treatments to cure AD, and early diagnosis of AD is crucial for timely therapeutic care to impede its progression. With the aging of the global population, a growth in the number of AD patients is expected, moreover, it was predicted that the number of patients suffering from AD will triple in 2050 [2]. Therefore, the computer-aided diagnosis of AD has attracted considerable interest in recent years.

AD-related biomarkers are used for its diagnosis, such as magnetic resonance imaging (MRI) features that contain information on the atrophy pattern of neurodegeneration [3], fluorodeoxyglucose positron emission tomography (FDG-PET) features that contain metabolic changes related to AD [4], and the concentrations of total tau (T-tau), phosphorylated tau (P-tau), and the 42-amino acid isoform of amyloid-β peptide (Aβ42) in the cerebrospinal fluid (CSF) [5]. In addition, the apolipoprotein E (APOE) ε4 gene is also an important risk factor of AD

☆ This work was supported by the National Natural Science Foundation of China [grant numbers 61901120, 61802065]; the Science and Technology Program of Fujian Province of China [grant number 2019YZ016006]; the Foundation of Educational and Scientific Research Projects for Young and Middle-aged Teachers of Fujian Province [grant numbers JAT200471, JAT190674]; and the High-level Talent Project of Xiamen University of Technology [grant number YKJ20013R].

* Corresponding author. College of Physics and Information Engineering, Fuzhou University, Fuzhou, 350116, China.
E-mail addresses: ttraveltong@gmail.com, ttraveltong@imperial-vision.com (T. Tong).

https://doi.org/10.1016/j.compbiomed.2021.104478
Received 28 January 2021; Received in revised form 5 May 2021; Accepted 5 May 2021
Available online 9 May 2021

[6]. Several studies have used these biomarkers to diagnose AD, which are binary classifications for AD in comparison with normal control (NC). However, there is also a transitional stage called mild cognitive impairment (MCI) between AD and NC. Therefore, multiclass diagnoses such as AD versus MCI versus NC are more practical.

In this study, we propose a framework that can diagnose AD using three-way classification (AD versus MCI versus NC) and four-way classification (AD versus progressive MCI (pMCI) versus stable MCI (sMCI) versus NC) with multimodal data including those related to MRI, positron emission tomography (PET), CSF, and genes. pMCI is defined as the MCI that will eventually convert to AD, and sMCI is the MCI that will not convert to AD. After performing the preprocesses of age correction, feature selection, and feature reduction, we implemented linear discriminant analysis (LDA) [7] for each modality and calculated a score to present the progress of AD using this modality. Then, a binary extreme learning machine (ELM)-based [8] tree decision strategy was utilized for multiclass diagnosis using these scores. Experiments were conducted using data from the AD Neuroimaging Initiative (ADNI) dataset, and the performance of the proposed framework improved when compared to a method using raw features. The contributions of this study are as follows:

i) LDA was implemented to calculate the scores from multimodal data, and these scores represent the information regarding the progress of AD in the corresponding modalities. In this way, the feature numbers were balanced for all modalities, which finally resulted in only one score, and the classifier could decide which modality was dominant without the influence of feature numbers.
ii) A binary ELM-based tree decision strategy was used for multiclass diagnosis, which demonstrated a superior performance when compared to support vector machine (SVM)-based multiclass classification.

The remainder of this paper is organized as follows. Section 2 briefly introduces the relevant studies. Section 3 presents the proposed method in detail. Section 4 presents the experimental data, settings, and results. In Section 5, we discuss the findings and limitations of this study. Finally, the conclusions of this study are presented in Section 6.

2. Related works

In previous studies, numerous methods have been proposed to discriminate AD from NC, and accuracies above 96% have been achieved [9,10] as there are apparent differences between these two groups, which makes it easy to discriminate between AD and NC. Certain studies have focused on predicting AD conversion from MCI by classifying pMCI and sMCI and have achieved accuracies above 80% [11,12]. However, the multiclass diagnosis of AD is a more practical but tougher task than binary classification tasks. In a comprehensive study [13], only a few algorithms have achieved an accuracy of over 60% in the three-way classification. Furthermore, there are also studies [14,15] that attempted a four-way diagnosis, which is more difficult than the three-way classification and is less accurate.

With different modalities of data reflecting AD-related pathological changes in different aspects, it is intuitive to fuse multimodal data to obtain complemented information and improve the performance of AD diagnosis. The simplest way to utilize multimodalities is to directly concatenate them. However, direct concatenation is not optimal, and there are better ways to fuse multimodal data, such as combining these data in a kernel space [10,16]. In recent years, deep learning techniques, such as convolutional neural networks (CNNs), have been employed to extract and fuse multimodal data from neuroimages, which can be used to improve the accuracy of AD diagnosis or prediction [17–19]. In the case of multiclass AD diagnosis, certain methods were proposed to fuse multimodal data and boost the accuracy. Tong et al. [20] employed a nonlinear graph fusion for four modalities in a three-way diagnosis of AD, and achieved an accuracy of 60.2%. Thung et al. [21] achieved a three-way diagnosis accuracy of 65.8% by using multitask deep learning with multimodal data of MRI, PET, and demographic information. Fang et al. [11] used a supervised Gaussian discriminative component analysis algorithm to delineate subtle changes between groups and achieved an accuracy of 66.29% for three-way classification. For four-way classification, Liu et al. [22] designed a zero-masking deep learning architecture for fusion of MRI and PET features and achieved an accuracy of 53.8%. Yao et al. [14] adopted MRI features combined with demographic and clinical data and achieved an accuracy of 54.4% by employing a relative importance-based feature selection and turning the four-way classification into five binary classification problems. Liu et al. [15] proposed a deep multitask multichannel CNN for joint classification and regression, and the accuracy of this four-way classification was 51.8%. However, there are biases in the feature numbers of modalities, which may lead to a bias in a modality with several features and hinder effective data fusion. The purpose of this study is to balance the feature numbers before fusing different modalities by using a scoring method with LDA so that the classifier can determine the weight of each modality and improve the efficiency of data fusion.

3. Methods

Four modalities were used in this study. The MRI was first analyzed using the FreeSurfer software [23]. FreeSurfer is a powerful tool for thickness calculation and whole brain segmentation and is usually used in AD analysis [11,24,25]. A total of 345 morphological features, including volume, surface area, and cortical thickness of cerebral regions, were provided by FreeSurfer. As 32 features were unavailable for most subjects, only 313 features were chosen as MRI features. For PET images, there are five regions that are considered to be predominantly related to AD, *i.e.,* the left angular, right angular, bilateral posterior cingulate, left inferior temporal, and right inferior temporal regions [26]. The mean, maximum, minimum, and standard deviation values of the intensities of these five regions were calculated from PET scans as features. Thus, 20 PET features were obtained. The other modalities included three CSF biomarkers, *i.e.*, levels of Aβ42, T-tau, and P-tau. Moreover, the genetic feature is a single variable indicating the presence of the APOE ε4 gene. Therefore, there were a total of 337 multimodal features for each subject. The overall framework of the proposed approach is illustrated in Fig. 1. First, the MRI features were preprocessed with age correction and feature selection for the following two reasons: the cerebral atrophy might not only be due to AD-related but also age-related factors, and not all MRI features were related to AD. Then, principal component analysis (PCA) [27] was implemented on MRI and PET to eliminate useless and interferential components. Then, all types of features (with CSF and genetic features combined as one) were individually processed using LDA [7], and only the most discriminative component was calculated as a score for each modality. Finally, three scores, together with the age, were input into the decision tree consisting of binary ELMs [8], which were pretrained with corresponding pairs of groups.

3.1. Preprocessing of magnetic resonance imaging features

The cerebral atrophy of AD patients is not only caused by AD, but also related to normal aging. Consequently, there might be age-related effects among MRI features. To reduce the interference of these effects, we first estimated the effects by fitting a linear regression model with MRI features from NC subjects [28]. Let us assume that there are N healthy subjects. $f_m \in R^{1\times N}$ is the vector of N values for the mth feature and $\psi \in R^{1\times N}$ is the vector of the ages of the subjects. When the linear model $f_m = a_m\psi + b_m$ was fitted, a_m can be regarded as an age-related effect for the mth feature. Then, we aligned the age effect to a target age for all subjects by calculating new features as $f'_m = a_m\,(C\text{-}\psi) + f_m$, where C is the target age all subjects were aligned.

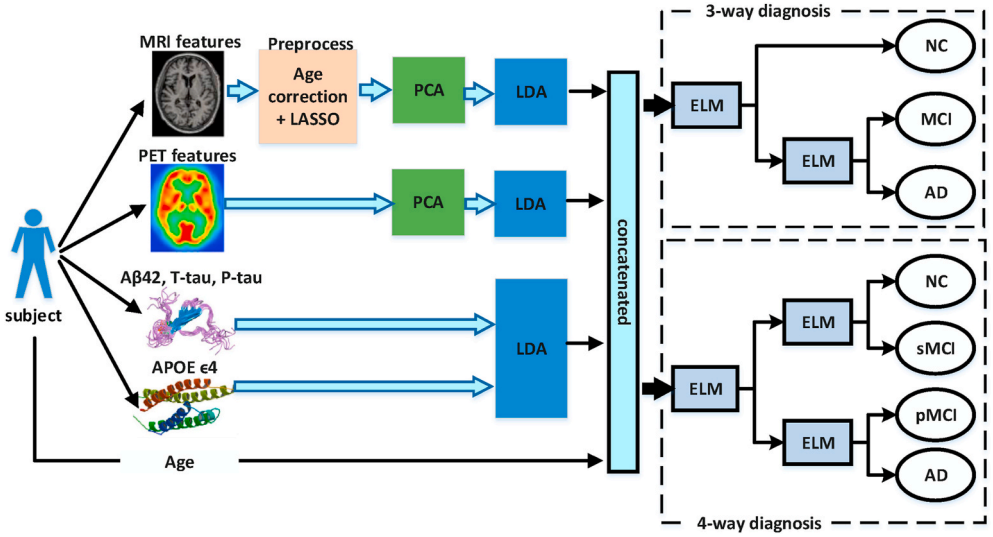


Fig. 1. Overall framework of the proposed approach.

As MRI features were calculated with global brain MRI analysis, features that are unrelated to AD were also present. To eliminate the interference of these unrelated features, we considered feature selection with the least absolute shrinkage and selection operator (LASSO) algorithm [29]. LASSO is an $L_{2,1}$-norm sparse regression model with the following formula:

$$\min_{\boldsymbol{\alpha}} 0.5\|\mathbf{y}-\mathbf{D}\boldsymbol{\alpha}\|_2^2+\lambda\|\boldsymbol{\alpha}\|_1 \tag{1}$$

where D is an N × M feature matrix that consists of N training samples with M features in each sample, and $y \in R^{N\times 1}$ is a vector of N labels. $\alpha \in R^{M\times 1}$ is the target sparse coefficient. Moreover, λ is a penalty coefficient that controls the sparseness of α, which was set to 0.015. After solving this model, only certain coefficients in α would be nonzero for the L_1-norm penalty, and these nonzero coefficients indicate that the corresponding features are important for predicting the labels. Therefore, the features with nonzero coefficients were selected as discriminative features, and the others were discarded. It should be noted that the training samples in D were from AD and NC subjects who had MRI data but lacked at least one other modality. Therefore, these samples were not present in the validation experiments. Thus, the double-dipping problem was avoided.

3.2. Features reduction with principal component analysis and modalities scored using linear discriminant analysis

PCA is an unsupervised learning method that uses an orthogonal transformation to convert correlated features into linearly uncorrelated features [27]. The correlated parts of the original features are fused into new features that appear as major components, and the remaining uncorrelated parts appear as interferential components. As there could be several correlations between MRI features and PET features, we implemented PCA on MRI and PET features. After PCA, the major components were retained, and the other components were discarded as interferential components.

The LDA-based scoring strategy is a key process in this approach. Different modalities are known to reflect different pathological changes in AD. When NC is converted to AD, the degree of pathological change in each modality can be assumed to progress along a path. The scoring strategy is to project the progressing path onto a line with LDA and calculate a score that represents the degree of pathological change. Using different modalities, we can obtain different scores representing different pathological changes. LDA is a supervised learning algorithm that can transform and project data, with the knowledge of class labels, onto a maximally discriminating vector, on which the projected data achieved maximized between-class variance and minimized within-class variance [7]. We individually implemented LDA on MRI, PET, and the other two modalities, and we adopted only one projection vector to calculate the value of the projected data as a score, as shown in Fig. 2. Finally, from LDA, we obtained scores representing AD-related pathological progress. There were three scores for all modalities, one from MRI, one from PET, and one from the combination of CSF and genetic biomarkers. These scores contained pathological information from different modalities and performed an equal role when they were combined for AD multiclass diagnosis.

3.3. Extreme learning machine-based decision tree

Although ELM can be used as a multiclass classifier directly [8], to discriminate different stages of AD more accurately, a decision tree strategy was implemented with binary ELMs. For three-way classification, the NC group was separated from the MCI and AD groups with one ELM first, and then, the MCI and AD groups were separated using another ELM. For four-way classification, the first ELM separated the NC and sMCI groups from the pMCI and AD groups, and then, two ELMs dealt with the classification of NC versus sMCI and pMCI versus AD. All

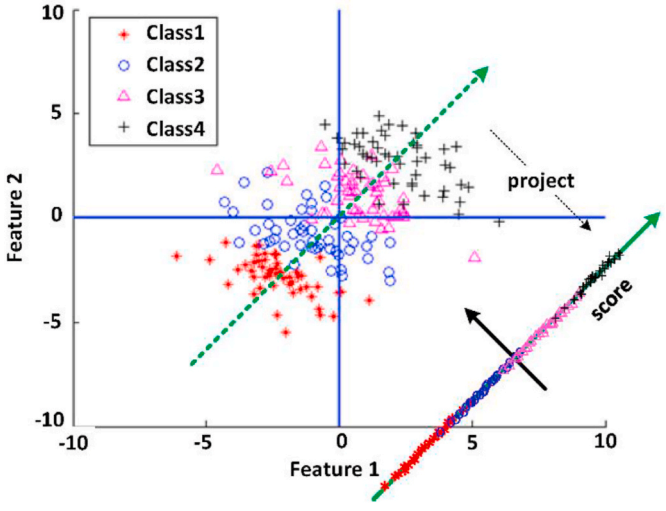


Fig. 2. The example of LDA scoring strategy.

ELMs were pretrained using the corresponding pairs of groups. Gaussian kernels [8] were adopted for the binary ELMs. Let us assume that we have N training samples [x_1, x_2, ⋯, x_N] and N labels, and x_n represents the vector of the nth sample consisting of M features. $Y \in R^{N\times G}$ is a ground truth label matrix with N rows. In each row, there are G elements corresponding to G groups, and the element of the true label is set to 1, while the others are set to −1. When we obtain a new sample x, the label of x can be predicted as

$$f(x) = \begin{bmatrix} K(\boldsymbol{x}, \boldsymbol{x}_1) \\ K(\boldsymbol{x}, \boldsymbol{x}_2) \\ \vdots \\ K(\boldsymbol{x}, \boldsymbol{x}_N) \end{bmatrix} (\boldsymbol{\Omega} + \boldsymbol{I}/C)^{-1}\boldsymbol{Y} \tag{2}$$

where K (x, x_n) is the Gaussian kernel described as

$$K(\boldsymbol{u}, \boldsymbol{v}) = \exp\left(-\|\boldsymbol{u} - \boldsymbol{v}\|^2/\gamma\right) \tag{3}$$

and Ω is an N × N kernel matrix that is related only to the training samples, which can be calculated in the training phase as

$$\boldsymbol{\Omega} = \begin{bmatrix} K(\mathbf{x}_1, \mathbf{x}_1) & & K(\mathbf{x}_1, \mathbf{x}_N) \\ K(\mathbf{x}_2, \mathbf{x}_1) & \cdots & K(\mathbf{x}_2, \mathbf{x}_N) \\ \vdots & & \vdots \\ K(\mathbf{x}_N, \mathbf{x}_1) & & K(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix} \tag{4}$$

The variable C is a regularization coefficient, which was set to 1. The variable γ is a parameter of the Gaussian kernel, which was set to 10 times the value of M.

4. Experiments and results

4.1. Data and experiment settings

The data used in this study were downloaded from the ADNI dataset, which recruited over 1800 participants with ages above 55. As not all modalities were available for all participants, only those subjects who had MRI, PET, CSF, and genetic data at the baseline time point were chosen. Thus, we obtained 200 NC, 441 MCI, and 105 AD subjects. It was determined that 110 MCI subjects were converted to AD during the following three years. Therefore, they were labeled as pMCI. A total of 208 MCI subjects remained as MCI and were labeled as sMCI. Moreover, 123 MCI subjects with unknown final status were excluded from the four-way diagnosis. The demographic and clinical information of these subjects are listed in Table 1, including sample counts, ages, mini-mental state examination (MMSE), and clinical dementia rating scale sum of boxes (CDR-SB). All these data were downloaded from the ADNI website. Specifically, the MRI, CSF, and genetic data were obtained from the data files of the TADPOLE Challenge, and the PET data were obtained from the UC Berkeley–FDG Analysis file.

In the cells of the second row, the first number is the total number with the numbers of females and males in brackets. SD is the standard deviation.

We considered the accuracy, recall, precision, and F1-score [11] to evaluate the performance of multiclass classification. The evaluation was conducted by five-fold cross-validation. 80% of the samples were used to train the LDA and classifiers, and the remaining 20% were used for testing. This trial was repeated five times with different 20% samples as the test set. To avoid sampling bias, we ran five-fold cross-validation 100 times with randomly permuted samples, and the mean and standard deviation of these 100 runs were calculated. In particular, as MCI is the obscurest group, the ratio of MCI would influence the evaluation. Therefore, only 150 MCIs were randomly selected from the MCI group in each run in the three-way classification to ensure that the ratio of MCI was one-third of all groups. For four-way classification, the ratio of sMCI and pMCI was approximately one-half of all groups. Therefore, all samples were used.

Table 1
The demographic information of subjects.

Mean ± SD	NC	AD	MCI	sMCI	pMCI
Count (F/M)	200 (93/107)	105 (36/69)	441 (191/250)	208 (92/116)	110 (47/63)
Age	73.9 ± 6.0	75.7 ± 8.0	72.4 ± 7.4	71.8 ± 7.1	73.9 ± 7.2
CDR-SB	0.0 ± 0.1	4.4 ± 1.6	1.4 ± 0.9	1.2 ± 0.6	1.9 ± 1.0
MMSE	29.0 ± 1.2	23.2 ± 2.0	27.9 ± 1.7	28.1 ± 1.7	27.1 ± 1.7

All the experiments were conducted using Python 3.6.4. The LASSO, PCA, and LDA algorithms were implemented using the Lasso, PCA, and LDA modules in the scikit-learn package. LASSO was trained with the NC and AD samples, which were not included in the subsequent experiments. PCA was trained with all samples without knowing their labels, and the "n_components" parameter was set to 10 to retain top 10 major components. LDA was trained with 80% samples with their labels during five-fold cross-validation, and the "n_components" parameter was set to 1 to retain the most discriminative component as score. The target age of age correction was set to 75, which is near the average age (73.3) of all subjects.

4.2. Performance of proposed method

To demonstrate the performance of the proposed method on multiclass diagnosis, we compared it with the original method, in which the raw features were concatenated directly and classified using SVM. The results are shown in Fig. 3, from which we can observe that the proposed method can significantly improve the performance in terms of recall, precision, F1-score, and accuracy. In particular, the proposed method increased the accuracy by up to 66.7% in the three-way diagnosis, which is 11.6% higher than that of the original method, and by up to 57.3% in the four-way diagnosis, which is also 11.6% higher than that of the original method. We can also identify that the low performance of multiclass diagnosis is primarily due to the low values of recall and precision of MCI or sMCI and pMCI. Therefore we also show the confusion matrix in Fig. 4. We determined that, in the original method, all groups were likely to be misdiagnosed with neighboring groups, and the proposed method can reduce the ratio of misdiagnosis.

We conducted experiments to reveal the impacts of all processes by adding different factors step by step. In each step, new factor was added to the framework of previous step, and 100 times 5-fold cross validation were repeated to calculate the performances of each step. The results are shown in Fig. 5. From these results, we determined that LDA had the most significant impact on the F1-score and accuracy, indicating the importance of LDA for this method. These results also demonstrated that the contributions of LASSO and PCA, and the ELM-based tree decision strategy, instead of SVM, can also improve the accuracy and F1-score. We also conducted experiments in different settings to thoroughly investigate the proposed method, and the results are listed in Table 2. ELM can be used as a multiclass classifier. However, the binary ELM-based tree decision can achieve a better performance, as shown in the first two rows of Table 2. The third to tenth rows show the impact of different modalities, and we determined that the MRI has the most important effect on the performance, followed by CSF and PET. Moreover, the genetic feature is the smallest data, but the performance also declined by approximately 1% in the four-way diagnosis without the gene information. In the last four rows, we compared the performance with different cross-validations. It can be observed that the performance in ten-, seven- and three-fold cross-validations were similar with the default (five-fold cross-validation). In contrast, a marginal decline is observed in the two-fold cross-validation, which indicates the robustness of the proposed method. In Fig. 6, we investigated the effect of age correction: when age correction was used in alignment to different ages from 65 to 85, it had similar performance with the default setting (aligned to 75 years).

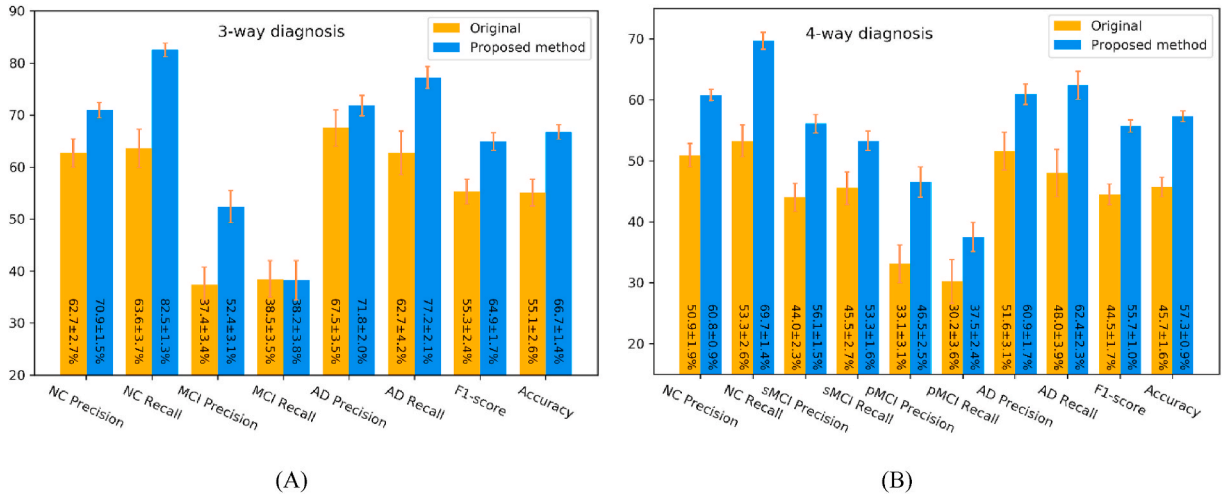


Fig. 3. Performance of the proposed method. (A) Performance of 3-way diagnosis. (B) Performance of 4-way diagnosis.

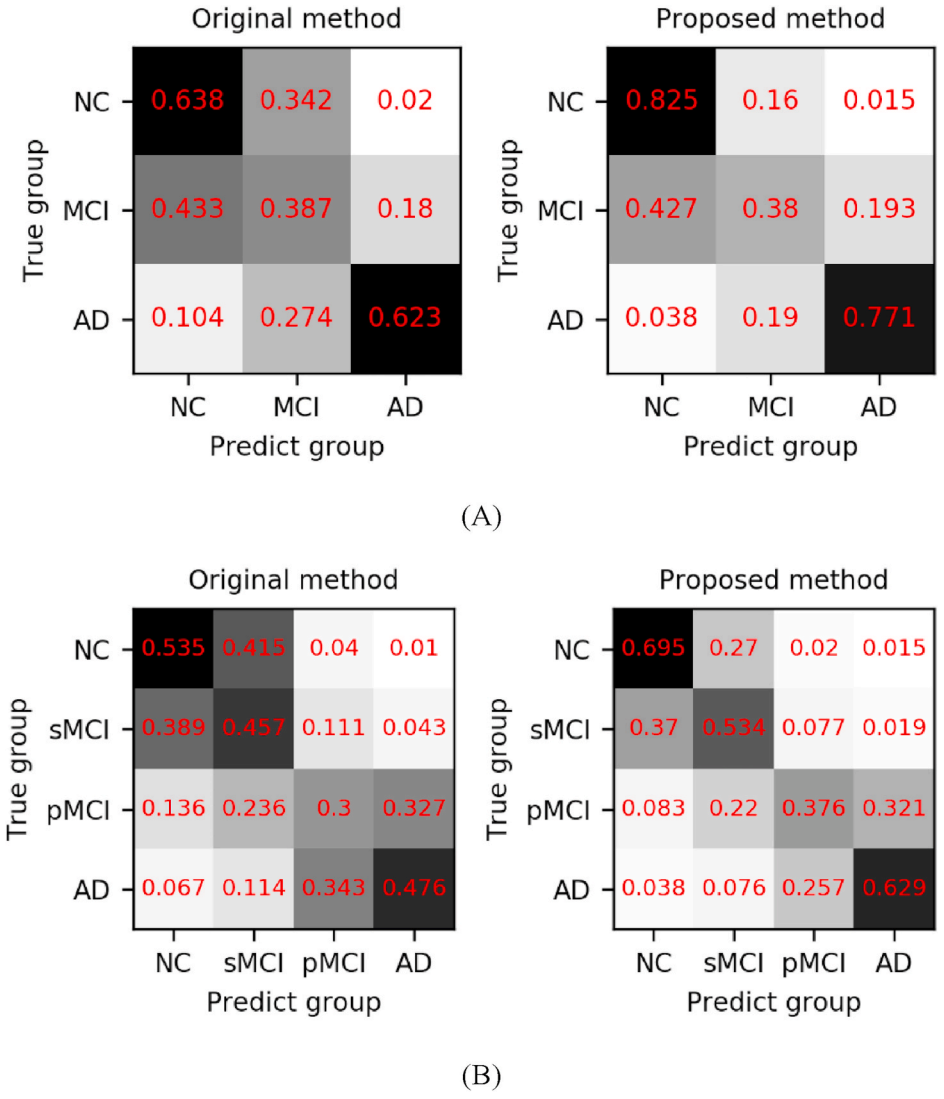
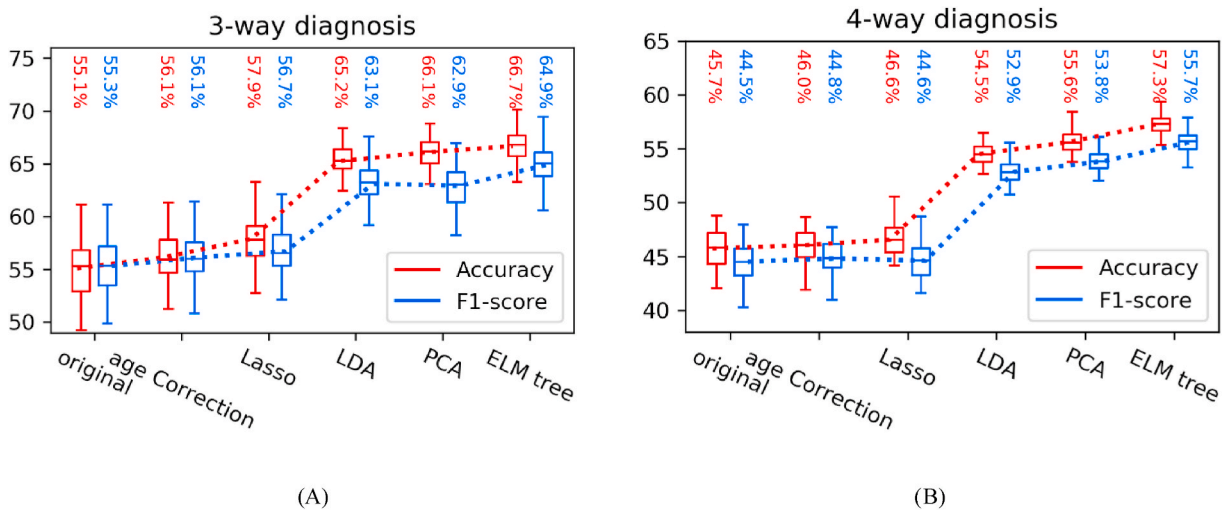


Fig. 4. Confusion matrixes. The red numbers are ratios. (A) Confusion matrixes of 3-way diagnosis. (B) Confusion matrixes of 4-way diagnosis.



(A) (B)

Fig. 5. Performance of multiclass diagnosis by adding different factors step by step.

Table 2
Performance in different settings.

Different settings	3-way diagnosis		4-way diagnosis (%)	
	Accuracy (%)	F1-score (%)	Accuracy (%)	F1-score (%)
Proposed method	66.7 ± 1.4	64.9 ± 1.7	57.3 ± 0.9	55.7 ± 1.0
ELM multiclass	65.5 ± 1.2	61.6 ± 1.7	55.5 ± 0.8	52.9 ± 0.9
Only MRI	61.5 ± 1.2	59.0 ± 1.4	49.6 ± 0.7	46.8 ± 0.8
Only PET	56.3 ± 1.2	54.8 ± 1.4	43.5 ± 0.9	43.1 ± 0.8
Only CSF	55.0 ± 1.0	53.1 ± 1.3	40.3 ± 0.8	37.0 ± 1.0
Without MRI	62.5 ± 1.5	61.5 ± 1.7	50.6 ± 0.8	49.6 ± 0.8
Without PET	65.5 ± 1.2	63.6 ± 1.3	55.4 ± 0.8	53.3 ± 0.9
Without CSF	63.9 ± 1.5	62.5 ± 1.7	54.6 ± 0.9	53.3 ± 1.0
Without Gene	66.2 ± 1.4	64.3 ± 1.6	56.4 ± 0.9	54.8 ± 1.0
Without Age	65.5 ± 1.2	63.3 ± 1.6	53.1 ± 1.0	51.9 ± 1.1
10-fold CV	66.5 ± 1.3	64.7 ± 1.5	57.4 ± 0.7	55.7 ± 0.8
7-fold CV	66.6 ± 1.3	64.9 ± 1.6	57.1 ± 0.7	55.4 ± 0.8
3-fold CV	66.3 ± 1.5	64.7 ± 1.7	56.8 ± 1.1	55.3 ± 1.3
2-fold CV	65.9 ± 1.6	64.4 ± 1.8	55.8 ± 1.6	54.4 ± 1.6

Note: In cells, the two number represent the mean and standard deviation of 100 runs. CV is the abbreviation of cross validation.

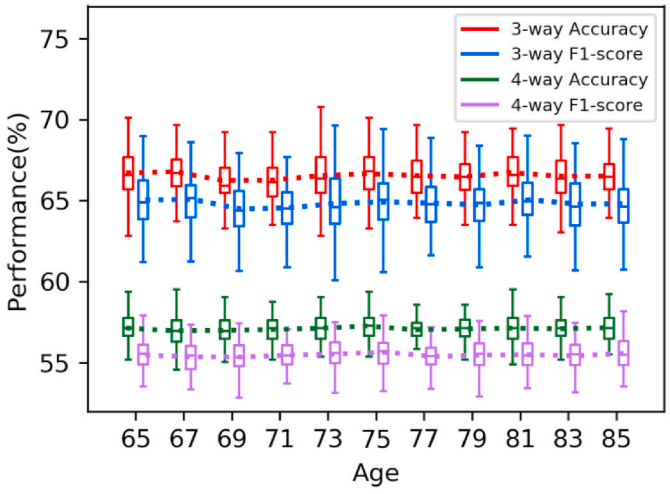


Fig. 6. Age correction with aligning to different ages.

Table 3
Comparison with other methods.

Methods	validation	3-way accuracy	4-way accuracy
Nonlinear graph fusion [20]	4-fold CV	60.2%	–
Age-dependent z-score + LDA [24]	10-fold CV	63.0%	–
Multi-task deep learning [21]	10-fold CV	65.8%	–
Gaussian discriminative component analysis [11]	9-fold CV	66.29%	53.9%
Zero-masking deep learning architecture [22]		–	53.8%
Feature selection + Ensemble learning [14]	10-fold CV	–	54.4%
Deep multi-task multi-channel learning [15]	Fixed dataset		51.8%
This study	10-fold CV	66.5%	57.4%
This study	5-fold CV	66.7%	57.3%

4.2.1. *Comparison with other methods*

We also compared the proposed method with other studies pertaining to multiclass AD diagnosis, and the results are listed in Table 3. Among these methods, the proposed method showed promising performance in both the three- and four-way classifications. In recent years, CNN-based approaches have been proposed for AD detection, but most of them have been used in binary classification. To compare with these methods, we also conducted experiments in binary classification (using a binary ELM) of NC versus AD and sMCI versus pMCI. The comparison results are listed in Table 4. From these results, we can observe that the CNN-based approaches achieved superior performances in NC versus AD classification, which is better than the proposed method. However, in the classification of neighboring groups, such as sMCI versus pMCI, the proposed method demonstrated better accuracy.

Table 4
Comparison with CNN based approaches in binary classification.

Methods	validation	NC vs. AD	sMCI vs. pMCI
hybrid CNN and RNN [30]	5-fold CV	89.1%	72.5%
Ensemble of deep CNN [19]	4-fold CV	99.3%	–
3D-CNN and FSBi-LSTM [17]	10-fold CV	94.8%	65.4%
Multi-Modality 3D CNN [18]	Fixed dataset	90.1%	76.9%
3D densely connected CNN [31]	Fixed dataset	97.4%	78.8%
CNN and ensemble learning [32]	5-fold CV	84%	62%
This study	5-fold CV	93.4%	81.2%

5. Discussion

In this study, we applied an LDA-based scoring strategy for multimodal data fusion for the task of multiclass AD diagnosis. With the scores that were analyzed using LDA and other processes, such as age correction, LASSO, PCA, and ELM-based decision tree classifier, the proposed method achieved a superior performance than the original method for both three- and four-way diagnoses. The experimental results demonstrated that the LDA-based scoring can significantly improve the performance of multiclass diagnosis, indicating the important role of LDA in this method.

LDA-based scoring is the key process in the proposed method. It can reduce the features of each modality into a single score. Different modalities reflect different aspects of the pathological changes of AD, such as the MRI reflects the cerebral atrophy and PET reflects the abnormalities in cerebral glucose metabolism. The information regarding the different pathological changes was contained in the features of different modalities. Therefore, directly concatenating these features is not an efficient method. Moreover, different numbers of modality features would lead to bias in classification. Therefore, it is reasonable to calculate scores representing different AD-related pathological changes using LDA, resulting in a more efficient the fusion of multimodal data. Feature reduction was primarily performed by LDA, although LASSO and PCA can also reduce features, the purpose of LASSO and PCA processes was to eliminate useless and interferential components and help LDA to calculate more precise scores.

Multiclass AD diagnosis is a challenging task. The performance of multiclass diagnosis is significantly lower than that of binary diagnosis for AD/NC [20], because there are several similarities between MCI and AD or between MCI and NC. The inclusion of MCI would lead to a high probability of misdiagnosis between MCI and NC or between MCI and AD. We have shown the confusion matrix of the three- and four-way classifications in Fig. 4, in which we can observe that many misdiagnoses occurred between neighboring groups. With the proposed method, the ratios of misdiagnosis between neighboring groups were reduced. However, there were still several MCI subjects who were misdiagnosed as other groups. These results indicate that the precise classification of MCI or sMCI and pMCI is the key to improving the accuracy of AD multiclass diagnosis, and our future work will study a more precise method for the discrimination of MCI from AD and NC.

The comparison results in Table 4 indicate that the CNN is a powerful technique in image analysis, and the result show a superior performance in NC versus AD classification. However, the performance was not ideal when discriminating between sMCI and pMCI. The reason for this might be that the difference between NC and AD is obvious, and it is easy for CNN to learn the AD-related pathological information from the training phase. However, the difference between sMCI and pMCI or between neighboring groups are subtle and inter-subject variables can easily result in interference, which hinders the efficiency of the CNN. Therefore, we chose the extracted features from MRI and PET rather than CNN for AD multiclass diagnosis in this study. In the future, for the utilization of CNN in multiclass diagnosis, we assume that whole brain segmentation should be considered, which can extract morphological features for multiclass classification.

Cognitive scores were not included in this study, although their inclusion can achieve a higher accuracy [33]. As cognitive scores were directly used in the clinical diagnosis, especially the labels of ADNI were primarily dependent on MMSE and CDR-SB, we thought the results would be biased and overestimated if the cognitive scores were used. Therefore, we excluded the use of cognitive scores in this study.

6. Conclusion

In this study, we proposed an LDA-based scoring strategy approach for AD multiclass diagnosis in the presence of four modalities, *i.e.*, MRI, FDG-PET, CSF, and genetic features. The LDA was used to calculate a score representing the pathological information from each modality, and the scores from different modalities ensured that the classifier could easily discriminate between different groups. LASSO and PCA were used to exclude irrelevant and interferential components before LDA, and a binary ELM-based tree decision classifier was built for multiclass classification. The experimental results indicated that the LDA scoring significantly improved the multiclass diagnosis. Benefiting from the information obtained from multiple modalities and the scoring strategy, we achieved a promising performance with an accuracy of 66.7% and F1-score of 64.9% for three-way diagnosis, and an accuracy of 57.3% and F1-score of 55.7% for four-way diagnosis, which were significantly better than the original method. When compared to other studies, the proposed approach also showed a better performance. Although multimodal data help to improve the performance of AD diagnosis, the requirement of too many modalities would limit the practical usage of this approach. However, the more efficient multimodal fusion approach in this study is still useful for further AD studies, such as AD longitudinal trajectory modeling.

Declaration of competing interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in the manuscript entitled.

References

[1] A. Ott, et al., Prevalence of Alzheimer's disease and vascular dementia: association with education. The Rotterdam study, Br. Med. J. 310 (6985) (1995) 970–973.
[2] T. Vos, et al., Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015, Lancet 388 (10053) (Oct. 2016) 1545–1602.
[3] Q. Feng, Z.X. Ding, MRI radiomics classification and prediction in Alzheimer's disease and mild cognitive impairment: a review, Current Alzheimer Res. Rev. 17 (3) (2020) 297–309.
[4] A. Drzezga, et al., Diagnostic utility of 18F-Fluorodeoxyglucose positron emission tomography (FDG-PET) in asymptomatic subjects at increased risk for Alzheimer's disease, European J. Nuclear Med. Mol. Imag. Rev. 45 (9) (Jul 2018) 1487–1496.
[5] E. Niemantsverdriet, S. Valckx, M. Bjerke, S. Engelborghs, Alzheimer's disease CSF biomarkers: clinical indications and rational use, Acta Neurol. Belg. 117 (3) (Sep. 2017) 591–602.
[6] M. Vounou, et al., Sparse reduced-rank regression detects genetic associations with voxel-wise longitudinal phenotypes in Alzheimer's disease, Neuroimage 60 (1) (Mar. 2012) 700–716.
[7] A. Tharwat, T. Gaber, A. Ibrahim, A.E. Hassanien, Linear discriminant analysis: a detailed tutorial, AI Commun. 30 (2) (2017) 169–190.
[8] G.B. Huang, H. Zhou, X. Ding, R. Zhang, Extreme learning machine for regression and multiclass classification, IEEE Trans. Syst. Man Cybernet. 42 (2) (Apr. 2012) 513–529.
[9] J. Kim, B. Lee, Identification of Alzheimer's disease and mild cognitive impairment using multimodal sparse hierarchical extreme learning machine, Hum. Brain Mapp. 39 (9) (Sep. 2018) 3728–3741.
[10] Q. Li, X. Wu, L.L. Xu, K.W. Chen, L. Yao, I. Alzheimers, Dis neuroimaging, "classification of Alzheimer's disease, mild cognitive impairment, and cognitively unimpaired individuals using multi-feature kernel discriminant dictionary learning, Front. Comput. Neurosci. 11 (Jan 2018) 14.
[11] C. Fang, et al., Gaussian discriminative component analysis for early detection of Alzheimer's disease: a supervised dimensionality reduction algorithm, J. Neurosci. Methods 344 (Oct 2020) 11.
[12] W.M. Lin, et al., Predicting Alzheimer's disease conversion from mild cognitive impairment using an extreme learning machine-based grading method with multimodal data, Front. Aging Neurosci. 12 (9) (Apr 2020).
[13] E.E. Bron, et al., Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: the CADDementia challenge, Neuroimage 111 (2015) 562–579.
[14] D. Yao, V.D. Calhoun, Z. Fu, Y. Du, J. Sui, An ensemble learning system for a 4-way classification of Alzheimer's disease and mild cognitive impairment, J. Neurosci. Methods 302 (2018) 75–81.
[15] M.X. Liu, J. Zhang, E. Adeli, D.G. Shen, Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis, IEEE Trans. Biomed. Eng. 66 (5) (May 2019) 1195–1206.
[16] J. Young, M. Modat, M.J. Cardoso, A. Mendelson, D. Cash, S. Ourselin, Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in

patients with mild cognitive impairment, Neuroimag. Clin. 2 (1) (May 2013) 735–745.
[17] C.Y. Feng, et al., Deep learning framework for Alzheimer's disease diagnosis via 3D-CNN and FSBi-LSTM, IEEE Acc. 7 (2019) 63605–63618.
[18] Y.C. Huang, J.H. Xu, Y.C. Zhou, T. Tong, X.H. Zhuang, I. Alzheimers, Dis neuroimaging, "diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network, Front. Neurosci. 13 (May 2019) 12.
[19] X.S. Fang, Z.B. Liu, M.C. Xu, Ensemble of deep convolutional neural networks based multi-modality images for Alzheimer's disease diagnosis, IET Image Process. 14 (2) (Feb 2020) 318–326.
[20] T. Tong, K. Gray, Q. Gao, L. Chen, D. Rueckert, A.s.D.N. Initiative, Multi-modal classification of Alzheimer's disease using nonlinear graph fusion, Pattern Recogn. 63 (Mar. 2017) 171–181.
[21] K.-H. Thung, P.-T. Yap, D. Shen, Multi-stage diagnosis of Alzheimer's disease with incomplete multimodal data via multi-task deep learning, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, 2017, pp. 160–168.
[22] S. Liu, et al., Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 62 (4) (Nov. 2015) 1132–1140.
[23] B. Fischl, et al., Automatically parcellating the human cerebral cortex, Cerebr. Cortex 14 (1) (Jan. 2004) 11–22.
[24] L. Sørensen, et al., Differential diagnosis of mild cognitive impairment and Alzheimer's disease using structural MRI cortical thickness, hippocampal shape, hippocampal texture, and volumetry*, Neuroimag. Clin. 13 (2017) 470–482.
[25] A. Bartos, D. Gregus, I. Ibrahim, J. Tintera, Brain volumes and their ratios in Alzheimer's disease on magnetic resonance imaging segmented using Freesurfer 6.0, Psychiatry Res. Neuroimaging. 287 (May 2019) 70–74.
[26] S.M. Landau, et al., Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI, Neurobiol. Aging 32 (7) (Jul. 2011) 1207–1218.
[27] S.L. Sell, S.G. Widen, D.S. Prough, H.L. Hellmich, Principal component analysis of blood microRNA datasets facilitates diagnosis of diverse diseases, PloS One 15 (6) (Jun 2020) 26.
[28] J. Dukart, M.L. Schroeter, K. Mueller, A.s.D.N. Initiative, Age correction in dementia–matching to a healthy brain, PloS One 6 (7) (2011), e22193.
[29] S.L. Kukreja, J. Löfberg, M.J. Brenner, A least absolute shrinkage and selection operator (LASSO) for nonlinear system identification, IFAC Proc. Vol. 39 (1) (Jan. 2006) 814–819.
[30] F. Li, M.H. Liu, I. Alzheimer's Dis Neuroimaging, A hybrid convolutional and recurrent neural network for Hippocampus analysis in Alzheimer's disease, J. Neurosci. Methods 323 (Jul 2019) 108–118.
[31] J. Zhang, B. Zheng, A. Gao, X. Feng, D. Liang, X. Long, A 3D densely connected convolution neural network with connection-wise attention mechanism for Alzheimer's disease classification, Magn. Reson. Imag. 78 (2021) 119–126.
[32] D. Pan, et al., Early detection of Alzheimer's disease using magnetic resonance imaging: a novel approach combining convolutional neural networks and ensemble learning, Front. Neurosci. 14 (May 2020) 19.
[33] T. Altaf, S.M. Anwar, N. Gul, M.N. Majeed, M. Majid, Multi-class Alzheimer's disease classification using image and clinical features, Biomed. Signal Process Contr. 43 (2018) 64–74.