# Early Alzheimer's disease diagnosis with the contrastive loss using paired structural MRIs

Hezhe Qiao [a,b,1], Lin Chen [a,1], Zi Ye [c], Fan Zhu [a,*]

[a] *Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, China*
[b] *University of Chinese Academy of Sciences, BeiJing 100049, China*
[c] *Johns Hopkins University, Baltimore, MD 21218, United States of America*

A B S T R A C T

*Background and objective:* Alzheimer's Disease (AD) is a chronic and fatal neurodegenerative disease with progressive impairment of memory. Brain structural magnetic resonance imaging (sMRI) has been widely applied as important biomarkers of AD. Various machine learning approaches, especially deep learning-based models, have been proposed for the early diagnosis of AD and monitoring the disease progression on sMRI data. However, the requirement for a large number of training images still hinders the extensive usage of AD diagnosis. In addition, due to the similarities in human whole-brain structure, finding the subtle brain changes is essential to extract discriminative features from limited sMRI data effectively.

*Methods:* In this work, we proposed two types of contrastive losses with paired sMRIs to promote the diagnostic performance using group categories (G-CAT) and varying subject mini-mental state examination (S-MMSE) information, respectively. Specifically, G-CAT contrastive loss layer was used to learn the closer feature representation from sMRIs with the same categories, while ranking information from S-MMSE assists the model to explore subtle changes between individuals.

*Results:* The model was trained on ADNI-1. Comparison with baseline methods was performed on MIRIAD and ADNI-2. For the classification task on MIRIAD, S-MMSE achieves 93.5% of accuracy, 96.6% of sensitivity, and 94.9% of specificity, respectively. G-CAT and S-MMSE both reach remarkable performance in terms of classification sensitivity and specificity respectively. Comparing with state-of-the-art methods, we found this proposed method could achieve comparable results with other approaches.

*Conclusion:* The proposed model could extract discriminative features under whole-brain similarity. Extensive experiments also support the accuracy of this model, i.e., it provides better ability to identify uncertain samples, especially for the classification task of subjects with MMSE in 22–27. Source code is freely available at https://github.com/fengduqianhe/ADComparative.

© 2021 Elsevier B.V. All rights reserved.

## 1. Introduction

Alzheimer's Disease (AD) is a chronic and fatal neurodegenerative disease that can lead to the death of entire nerve cells and tissue loss, accounting for approximately 60% of all dementia cases [1]. There is currently no effective drug therapy to cure it, only clinical intervention in its development [2]. According to statistics, there are approximately 21–35 million AD patients worldwide, which causes huge social dementia-related costs. The main clinical manifestations of AD include progressive memory impairment,

cognitive impairment, personality changes, and language impairment [3]. The development of AD is a decades-long process and extremely difficult to diagnose in the early clinical stage. The affected subjects experience identifiable cognitive deficits without substantial dysfunction in the Mild Cognitive Impairment (MCI) stage, known as a prodromal stage of AD [4]. Due to the slow development of AD during the MCI stage and the earlier stages, it is very important to accurately identify AD and MCI patients for the early diagnosis of AD and further clinical interventions.

In recent years, structural magnetic resonance imaging (sMRI) has been widely used to find a morphometric pattern in this challenging task of computer-aided AD diagnosis [5]. sMRI scan could provide detailed information of internal anatomy and brain tissue morphology, such as white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF), which are beneficial for the identification
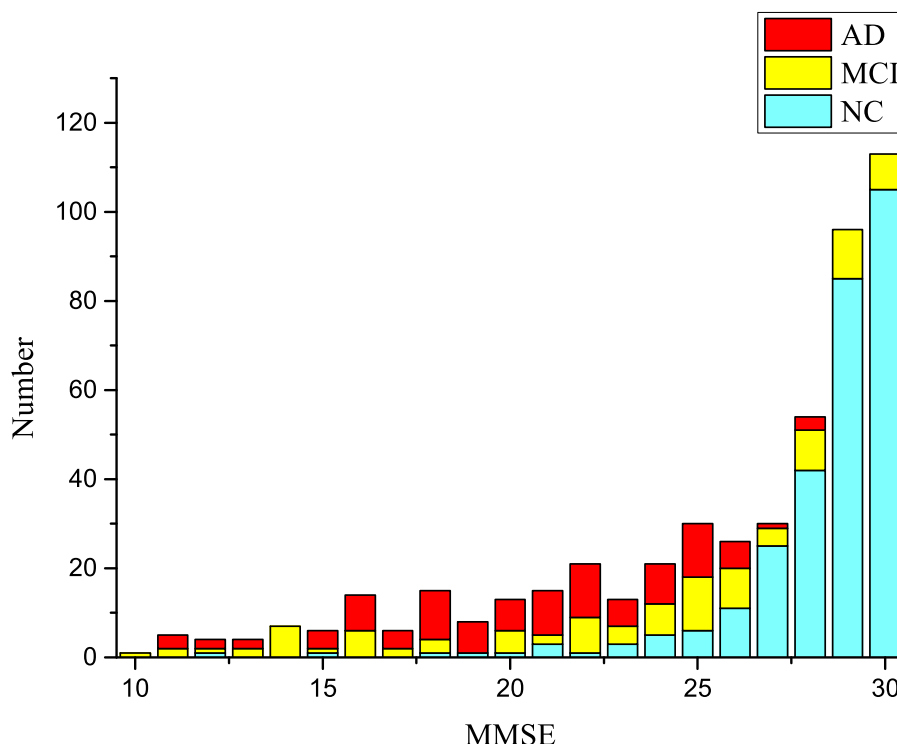
**Fig. 1.** Distribution of subjects' MMSE in three categories including NC, MCI, and AD. The MMSE of NC is mainly distributed in the range 27–30, while the MMSE of AD is mainly distributed in the range 10–25. The MMSE distribution of MCI has a wide range and also has a large overlap with the interval of AD and NC.

of abnormal structure brain changes. Additionally, as an important clinical indicator of AD that measures cognitive state for potential patient, mini-mental state examination(MMSE) is also widely used in clinical and research settings to measure the cognitive status of dementia.

Subjects in AD research can be divided into three stages, Normal Control (NC), MCI, and AD. Since the development of AD is very hidden, accurately identifying the stage of the subject is an important and challenging task for AD diagnosis. As Fig. 1 shown is the MMSE distribution of more than 700 subjects at different stages. It can be clearly seen that the scores of each category are roughly distributed in different intervals, but there is a large overlap between these intervals, especially between MCI and NC. which also illustrates the challenge of the subjects' group classification tasks.

Some researchers have tried to use specific data to predict the progression of NC to AD and NC to MCI [6,7]. A large number of voxel-based morphology (VBM) research have achieved good results [8,9]. The traditional VBM methods mainly use computer-assisted systems to calculate the inherent characteristics of certain biomarkers, such as hippocampal volume, cortical disease, and subcortical volume [10]. On this basis, methods such as 3D texture analysis [11], biomarker measurements [12] are introduced for feature construction. These inherent features combined with clinical medical data and genetic data were used to build a classifier to predict the changing trends of subjects at different stages [13]. F. Previtali *et al.* proposed a feature extraction technique from patients' MRI brain scans [14]. Zhang *et al.* proposed a new multi-view clustering model called Consensus Multi-view Clustering (CMC) based on nonnegative matrix factorization for predicting the multiple stages of AD progression [15]. The method based on VBM is highly interpretable, but its realization requires professional software operation and a large amount of expert knowledge, where are easily affected by subjective individuals [16].

With the continuous development of deep learning technology, methods based on CNNs perform well on many tasks such as image classification, image segmentation [17,18]. Deep learning avoids manual feature extraction by constructing an end-to-end model, which is also widely used in sMRI feature extraction [19–21]. Numerous studies based on 3D CNNs at the subject level extract high-level features for AD diagnosis [22–24]. Since the hippocampus is a biomarker related to AD, Some researchers use the three-dimensional convolutional network to combine the global and local features of the hippocampus for early diagnosis of AD [25]. In addition to modeling at the subject level, the methods based on 3D-Patch and 2D-Slice combined with the voting machines have also achieved good results on the specific tasks [26,27]. Although these methods avoid the problem of data scarcity, it needs to train multiple classifiers at different positions leading to an increase in the number of parameters.

As ROI based [28] and landmark-based [29,30] methods can avoid the interference of invalid information, they are also widely used to creat more training samples in AD diagnosis. In addition to using sMRI alone, many studies based on deep learning have explored multi-modal data such as positron emission tomography (PET) and diffusion tensor imaging (DTI), demographic data[31], genetics, and cerebrospinal fluid biomarkers [32,33].

Although significant efforts have been devoted to the early diagnosis of AD using deep learning models, however, it is still an unidentified problem: (1) collecting AD, NC, and MCI samples are time-consuming and expensive, which leads to the lack of sufficient training samples to train complex deep learning models at the subject level. (2) A single case of sMRI data contain multiple tissues and have too much noise and irrelevant information, which could seriously affect the generalization ability of the training model with a limited number of training subjects. (3) As Fig. 2 shown, the sMRI anatomical abnormalities of AD and NC, MCI and NC are relatively small causing the existence of a lot of ambiguous
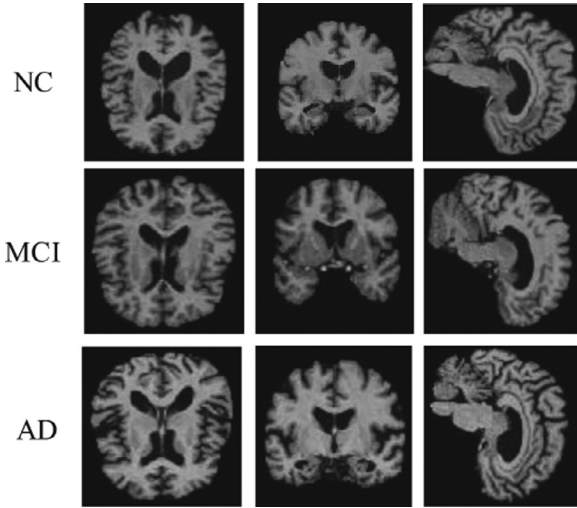
**Fig. 2.** Comparison of sMRI slices of NC, MCI and AD in the sagittal plane, coronal plane, and axial plane. All of sMRIs shown here have undergone pre-processing operations such as skull stripping.

samples. Considering all the above, it is necessary to explore subtle changes in disease progression from the limited sMRI data.

Considering most of the sMRIs of the whole-brain are similar, thus, simple feature extraction methods could not fully mine the subtle differences between subjects. In this work, we introduce a contrastive deep learning method to improve the AD prediction results by capturing the discriminative features. More specifically, group categories (G-CAT) contrastive loss and subject MMSE ranking (S-MMSE) loss, are proposed to measure the brain changes from the paired sMRIs. The contributions of our work are:

1. The group based contrastive loss layer is applied to restrict the feature representation of the paired images that belong to the same category to gather more closely.
2. S-MMSE ranking loss was added in the process of sMRI feature extraction to construct the ranking layer using paired sMRIs with varying MMSE values, which can assist the model to learn the subtle differences between individuals.
3. On the basis of the overall network structure of contrastive loss layers based on the paired sMRIs, the G-CAT and S-MMSE have good effects on both AD *vs*. NC, MCI *vs*. NC and AD *vs*. MCI *vs*. NC classification tasks and performed well than several states of the art methods.

## 2. Methods

In this section, first, we will introduce the overall architecture of the proposed model, and then describe the pair sMRI group categories contrastive loss layer and subject MMSE ranking loss layer respectively in detail.

### 2.1. Model structure overview

The overall architecture of our network is shown in Fig. 3. For 3D images, such as sMRI, we need to extend 2D convolution to 3D convolution that used for behavior recognition at the earliest. The form of 3D convolution operation is in the following:

$$u_j^l(x, y, z) = \sum_{\delta_x} \sum_{\delta_y} \sum_{\delta_z} F_k^{l-1}(x + \delta_x, y + \delta_y, z + \delta_z)$$
$$\times W_{kj}^l(\delta_x, \delta_y, \delta_z) \tag{1}$$

where $(x, y, z)$ is the coordinates of pixel in 3D image, $F_k^{l-1}$ is the $k$ feature map of the $l$ layer and $W_{kj}^l(\delta_x, \delta_y, \delta_z)$ is a three-dimensional

convolution kernel connecting the $k$th feature map of the $l-1$ layer to the $j$th feature map of the $l$ layer. $u_j^l(x, y, z)$ is the output of the convolutional layer, the new $j$th feature map of the $l$ layer. In order to achieve the comparison between samples, we use the siamese network whose input with two randomly selected sMRIs as our main network architecture. The respective attribute representations of the paired samples could be obtained through the neural network sharing the same weights [34].

Following the convolution layer is an AvgPooling layer which maps the multi-channel hidden layer to a vector.

For each subject in the paired input, the cross entropy loss $\mathcal{L}$ is defined as:

$$\mathcal{L} = -\frac{1}{C} \sum_{c=1}^{C} \frac{1}{N} \sum_{\mathbf{X}_n \in X} I\{y_n^c = c\} \log(P(y_n^c = c | \mathbf{X}_n : W)) \tag{2}$$

where $\mathcal{L}$ is the cross-entropy loss for classification and the $I\{\cdot\}$ is an indicator function. When $\{\cdot\}$ is true, $I\{\cdot\} = 1$, otherwise $I\{\cdot\} = 0$. $P(y_n^c = c | \mathbf{X}_n : W)$ is the probability of the subject $X_n$ being accurately classified into category $C$. For a pair of input sMRIs $X = (x_i, x_j)$, we construct a cross-entropy loss function $\mathcal{L}_{ij} = \mathcal{L}_i + \mathcal{L}_j$, where $\mathcal{L}_i$ and $\mathcal{L}_j$ denote the cross-entropy loss function of each sample in a pair MRI respectively.

### 2.2. Group category contrastive loss layer

Most patients with AD have the same pathological features on whole-brain sMRI, such as atrophy of the hippocampus and atrophy of the medial temporal lobe. In order to investigate both the differences and commonalities between groups, we first proposed the group categories (G-CAT) based contrastive loss layer model. As shown in the Fig. 4, the category factors $Cog_{ij}$ corresponding to paired sMRIs could be used as new ground truth to guide the model to learn more robust features. $Cog_{ij}$ is defined as the following:

$$Cog_{ij} = \begin{cases} 0 & y_i = y_j \\ 1 & y_i \neq y_j \end{cases} \tag{3}$$

where $y_i, y_j$ are the categories of inputs $x_i, x_j$ respectively. We construct the new feature information based on the difference between the paired MRI embedding feature, which could be used to distinguish whether the two input samples belong to the same category. When their categories are same, the $Cog_{ij}$ is 0, otherwise the $Cog_{ij}$ is 1. The hidden layer representation of paired sMRIs are $h_c^1, h_c^2$, We use the cascading of two vectors $h_c = [h_c^1, h_c^2]$ to represent the difference features between the input samples. The $p_{ij} = f(h_c)$ indicates the predicted probability that two subjects belong to the same category, where $f$ is a fully connected network mapping the embedding difference features into probabilities.

The loss function of G-CAT $\mathcal{L}^{CAT}$ is defined as:

$$\mathcal{L}_{ij}^{CAT} = -Cog_{ij} \log p_{ij} - (1 - Cog_{ij}) \log(1 - p_{ij}) \tag{4}$$

By incorporating this group comparative module, the new joint contrastive loss based subjects' groups $\mathcal{L}_{joint}^{CAT}$ is the following:

$$\mathcal{L}_{joint}^{CAT} = \sum_{m=i,j} \mathcal{L}_m + \alpha \sum_{m=i,j} \mathcal{L}_m^{CAT} \tag{5}$$

where $\alpha$ is a hyperparameter that adjusts the weights between the two loss functions.

### 2.3. Subject MMSE ranking loss layer

Compared to G-CAT, based on the assumption that subjects with similar scores may have the same categories, we design a subject MMSE (S-MMSE) ranking loss layer to measure the changes of various S-MMSE values. The structure of the ranking layer is
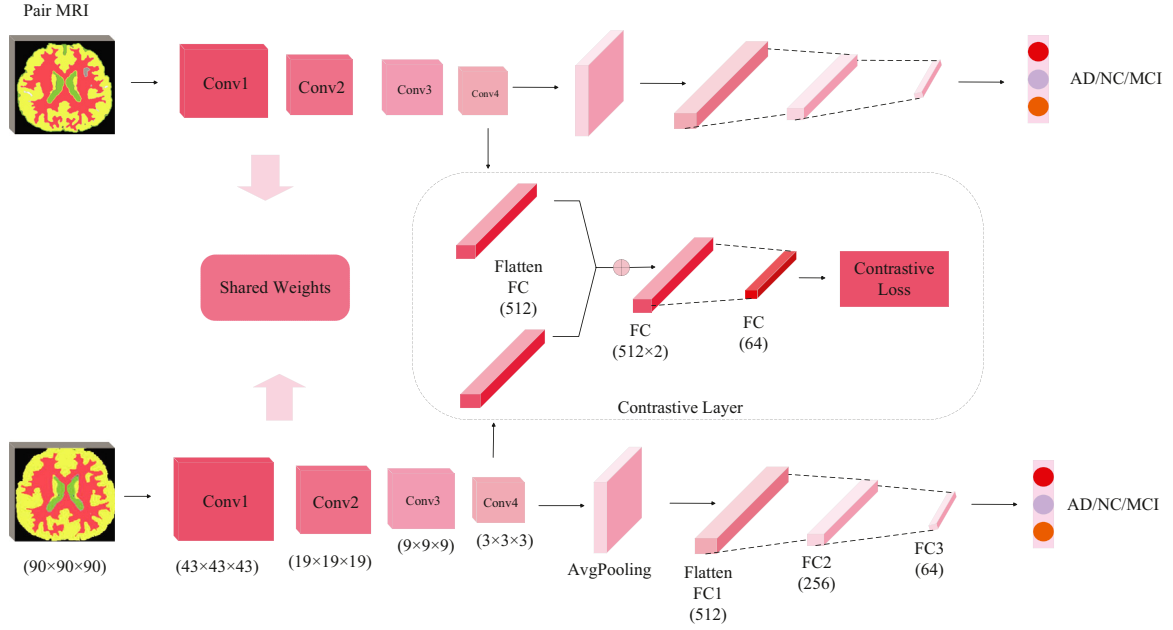
**Fig. 3.** The overall structure of our proposed model. The model includes the basic siamese network structure and the proposed contrastive loss layer module. The new feature representation was obtained by cascading two hidden layer vectors from two branches, and the final output of the model is the probability distribution of the corresponding category.
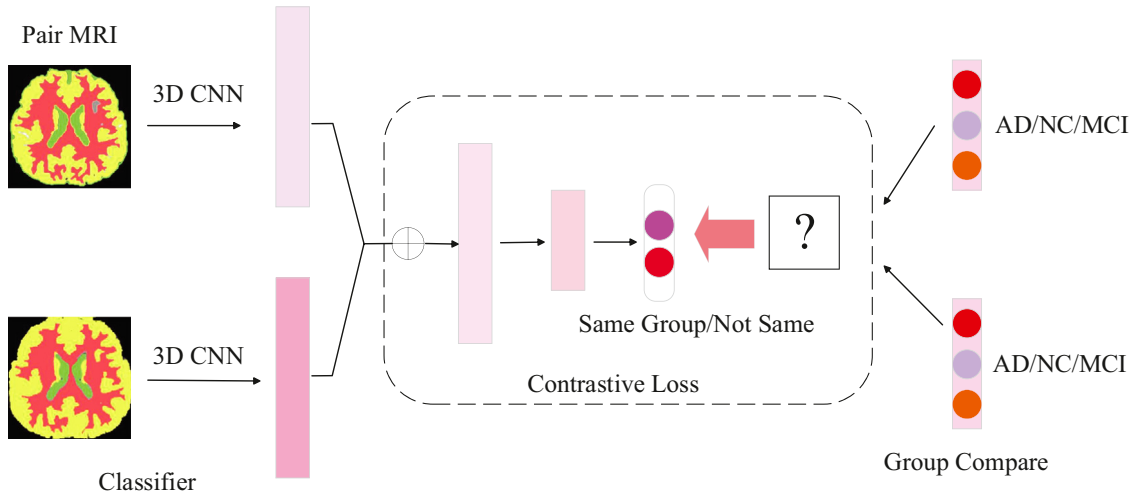


**Fig. 4.** Group categories contrastive loss layer. This module takes the paired sMRIs as input and extracts features through 3D CNNs with sharing weights. The obtained embedding layer was used to identify whether the input MRI groups are of the same category.

shown in Fig. 5. The distance between the MMSE of different subjects could provide more ranking information including equal, less, and greater. Similarly, given a pair of sMRI $m_i, m_j$ and the corresponding MMSE $s_i, s_j$. The ranking factor $Cos_{ij}$ is defined as:

$$Cos_{ij} = \begin{cases} 0 & s_j - s_i \leq \xi \\ 1 & 0 \leq |s_i - s_j| \leq \xi \\ 2 & s_i - s_j \geq \xi \end{cases} \quad (6)$$

where $\xi$ is a threshold that indicates similar subjects since the MMSE of the same categories could fluctuate up and down, but strictly equal. When $0 \leq |s_i - s_j| \leq \xi, \xi \geq 0$, we suppose there is no significant difference between the input sMRIs. It is noted that we did not consider the exact MMSE of each subject in the training, we only used the ranking information between the MMSE of different subjects, which could provide extra information for distinguishing the ambiguous samples. In the inference process for

AD diagnostic, sMRI will be the only input to our model without MMSE.

The ranking loss function $\mathcal{L}^{rank}$ is in the following:

$$\mathcal{L}_{ij}^{rank} = -Cos_{ij} \log p_{ij} - (1 - Cos_{ij}) \log(1 - p_{ij}) \quad (7)$$

where $p_{ij}$ is same as the definition in the G-CAT, but it represent the predicted probability of ranking type between the MMSE of pair subject. By incorporating this ranking layer module, the MMSE based joint contrastive loss function $\mathcal{L}_{joint}^{rank}$ is the designed as:

$$\mathcal{L}_{joint}^{rank} = \sum_{m=i,j} \mathcal{L}_m + \beta \sum_{m=i,j} \mathcal{L}_m^{rank} \quad (8)$$

where $\beta$ is a hyper-parameter that adjusts the weight between the two loss functions.
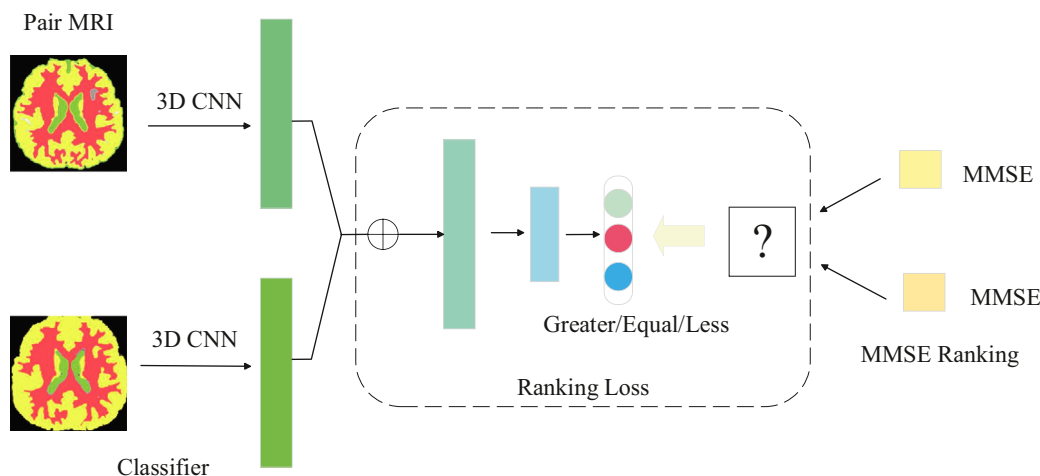
**Fig. 5.** Subject MMSE ranking loss layer. This module takes paired sMRI as inputs and extracts features through 3D CNNs with sharing weights. The obtained embedding layer was used to explore the MMSE ranking. The relationships of MMSE could be divided into three types including greater, less and equal.

**Table 1**
Demographic characteristics of the studied subjects from three different databases including ADNI-1, ADNI-2 and MIRIAD. (The values are denoted as mean ± standard deviation).

| Data | Diagnosis | Number | Age(Years) | Sex(M/F) | MMSE |
|------|-----------|--------|------------|----------|------|
| ADNI-1 | NC | 199 | 76.2±5.1 | 104/95 | 29.2±1.0 |
| | MCI | 332 | 74.9±7.2 | 206/126 | 26.9±2.0 |
| | AD | 141 | 75.6±7.6 | 74/67 | 23.1±2.5 |
| MIRIAD | NC | 23 | 70.4±7.2 | 12/11 | 29.4±1.2 |
| | AD | 46 | 70.0±7.1 | 19/27 | 20.5±5.7 |
| ADNI-2 | NC | 146 | 77.3±6.7 | 76/70 | 28.8±1.7 |
| | MCI | 114 | 79.3±7.1 | 74/40 | 23.6±6.0 |
| | AD | 111 | 75.6±7.8 | 67/44 | 21.9±3.8 |

### 2.4. Datasets and image preprocessing

#### 2.4.1. Datasets

The dataset used in this study was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) that is available at http://adni.loni.usc.edu/ [35]. Detailed information about MR acquisition procedures is available at the ADNI website. The ADNI data set contains two sub-datasets, including ADNI-1 and ADNI-2. The main goal of ADNI is to test whether serial sMRI, positron emission tomography (PET), other biomarkers, and clinical and neuropsychological evaluation can be combined to measure the progression of MCI and early AD. Subjects in the ADNI-1 have 1.5T T1-weighted structural sMRI data while ADNI-2 has 3.0T T1-weighted structural data. The ADNI-1 has 199 NC, 332 MCI, and 141 AD subjects. The ADNI-2 has 146 NC, 114 NC, and 111 AD subjects.

The MIRIAD is also a database of sMRI brain volume scans of Alzheimer's disease patients and healthy elderly people. Each participant was scanned multiple times at intervals of 2 weeks to 2 years. The purpose is to study the feasibility of using sMRI as an outcome indicator of clinical trials for the treatment of Alzheimer's disease [36]. The MIRIAD dataset has 1.5T T1-weighted structural MRI from 46 AD and 23 NC subjects. Note that in the MIRIAD, only MMSE score age and gender information are available. It can be noted that there are no MCI subjects in the MIRIAD data set. The details are included in Table 1.

#### 2.4.2. Image pre-processing

To explore the valuable information for the training model, all MRIs were prepocessed by a standard pipeline in CAT12 toolbox which is avaiable at http://dbm,neuro.uni-jena.de/cat/. Firstly, we perform anterior commissure (AC)-posterior commissure (PC) correction. In order to quickly perform AC-PC correction on a large number of MRI samples, we have adopted the steps based on Matlab and SPM12 toolbox to perform MRI correction in batches, which could be find at https://www.fil.ion.ucl.ac.uk/spm/software/spm12/. Then we registered the sMRIs from different datasets into the Montreal Neurological Institute (MNI) space. N3 algorithm was applied to correct the intensity inhomogeneity of all MRI. In addition, all the images were pre-processed by skull stripping and removed invalid areas, leaving only the brain locations. By removing the invalid region, the neural network can focus more on the feature extraction of the valid region and avoid introducing too much noise. The main steps of pre-processing is shown in the Fig. 6. Following is the min-max normalization to correct the intensity of those images. Finally, as an input of the neural network, all images need be resized to the same resolution. We chose image sizes as $90 \times 90 \times 90$ in our experiments, the impact of image sizes will be discussed in detail in the experiment.

### 2.5. Experimental settings

The proposed model is implemented on the Pytorch library on 2 NVIDIA GeForce GTX 1080Ti with 11G GPU memory. In this study, we have conducted three tasks, including AD *vs*. NC, MCI *vs*. NC and AD *vs*. MCI *vs*. NC classification. In order to avoid the influence of using models with different parameters, the batch size of all models is set to 12. The optimization method is Adam and we adjusted the learning rate to make each model converge to an optimal value. In order to avoid over-fitting, we also added an early stopping mechanism during the training process. We shuffled the original dataset to generate a new set of sMRI, which was combined with the original data for paired training. It is worth noting that the number of pair sMRIs is equal to the number of original sMRIs, we just randomly select a contrast sMRIs in the training set. In the ablation experiments section, we also discuss the impact of hyperparameters on our model. To validate and evaluate the model's generalization ability, we used the ADNI-1 dataset as the training dataset and evaluate the model on the MIRIAD and ADNI-2. For classification, the following measures
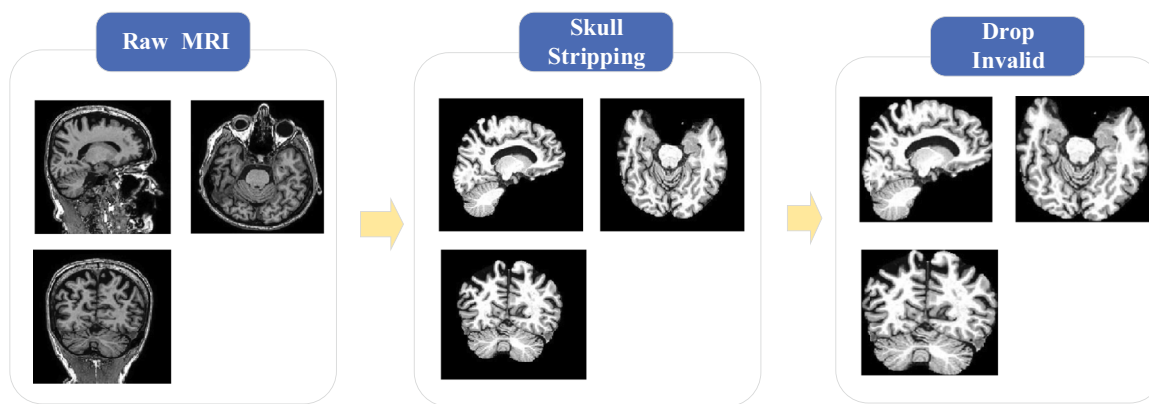
**Fig. 6.** MRIs are pre-processed by skull stripping and removed invalid areas.



(a) AD *vs.* NC on MIRIAD

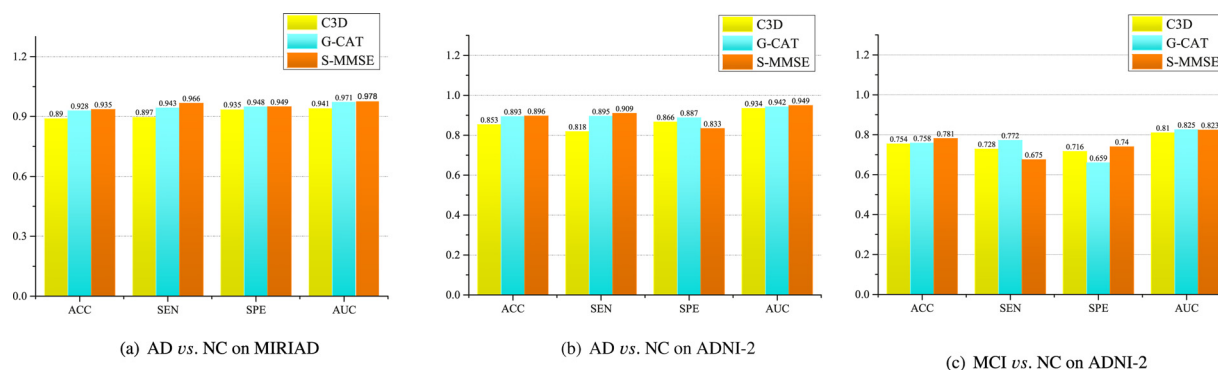(b) AD *vs.* NC on ADNI-2

(c) MCI *vs.* NC on ADNI-2

**Fig. 7.** Comparison of our proposed G-CAT, S-MMSE and the baseline modle C3D on four indicators including ACC, SEN, SPE and AUC. All the methods were trained on ADNI-1 and tested on MIRIAD and ADNI-2 respectively.

were computed for evaluation: classification accuracy (ACC), sensitivity (SEN), specificity (SPE), receiver operating characteristic sensitivity (ROC) curve, and under ROC curve (AUC).

## 3. Results

### 3.1. Comparison with baseline models

Due to the simplicity of structure and high generalization on small datasets, 3D neural networks (C3D) as a learning method for extracting spatio-temporal features can also perform feature extraction on 3D images [37], was used as a comparison method in this study. When $\alpha$ or $\beta$ in the loss function is set to 0, our proposed model degenerates and becomes a C3D model with two inputs. The comparison between C3D and our model can clear illustrate our superiority of adding the comparison module. The overall indicators comparison of models and ROC curve comparison are shown in the Figs. 7 and 8. We will discuss the effects of our proposed model on specific tasks on different data sets in detail.

### 3.1.1. Results on MIRIAD

Since there are only two categories of subjects in MIRIAD including AD and NC. In this experiment, we train the models on ADNI-1 and perform AD *vs.* NC classification on MIRIAD. The experimental results were shown in Table 2. The AUC of G-CAT and S-MMSE have achieved 97.1 and 97.8, respectively. Compared with the C3D model, G-CAT and S-MMSE perform well on all four indicators including ACC, SEN, SPE, and AUC, indicating that our proposed module can indeed extract more effective features.

We further compare the proposed group-based model and MMSE based model, we find the S-MMSE performed better than

**Table 2**
Result for AD *vs.* CN classification task with models trained on ADNI-1 and tested on MIRIAD.

| Method(%) | ACC | SEN | SPE | AUC |
|---|---|---|---|---|
| C3D | 89.0 | 89.7 | 93.5 | 94.1 |
| G-CAT | 92.8 | 94.3 | 94.8 | 97.1 |
| S-MMSE | **93.5** | **96.6** | **94.9** | **97.8** |

the G-CAT. It may be due to the significant difference in MMSE between AD and NC. The MMSE of AD is mainly distributed in the range of 20–25, while the score of NC is mainly distributed in the range of 28–30, which is conducive to the establishment of ranking information between sMRI data.

### 3.1.2. Results on ADNI-2

To better verify the effect of the model and the generalization performance of the proposed model, we also applied the model to the ADNI-2 for testing. Similar to ADNI-1, ADNI-2 contains three types of subjects, including AD, NC, and MCI. In this experiment, ADNI-1 was still used as a training dataset, the difference is that we add the MCI *vs.* NC classification task that is more difficult and more important in early AD diagnosis, because subjects in the MCI stage did not show obvious symptoms of dementia, leading to slight changes in sMRI compare with NC, e.g., the atrophy of the hippocampus, making it difficult for accurate identification.

The experimental results of two classificaiton tasks performed on ADNI-2 are shown in Table 3 and 4. For the AD *vs.* NC, G-CAT and S-MMSE both outperform the baseline. It is worth noting that S-MMSE still performs well than group-based models, which once again verified our conjecture that MMSE difference is beneficial for
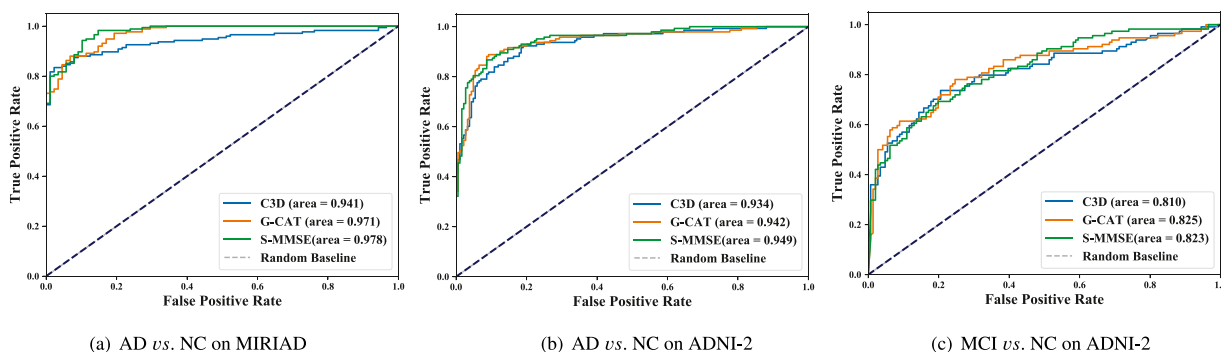
(a) AD *vs*. NC on MIRIAD          (b) AD *vs*. NC on ADNI-2          (c) MCI *vs*. NC on ADNI-2

**Fig. 8.** Comparison of the ROC curve between G-CAT, S-MMSE and the baseline modle C3D. All the methods were trained on ADNI-1 and tested on MIRIAD and ADNI-2 respectively.

**Table 3**
Results for AD *vs*. CN and MCI *vs*. NC classification tasks with models trained on ADNI-1 and tested on ADNI-2.

| Method(%) | AD *vs*. NC | | | | MCI *vs*. NC | | | |
|---|---|---|---|---|---|---|---|---|
| | ACC | SEN | SPE | AUC | ACC | SEN | SPE | AUC |
| C3D | 85.3 | 81.8 | 86.6 | 93.4 | 75.4 | 72.8 | 71.6 | 81.0 |
| G-CAT | 89.3 | 89.5 | **88.7** | 94.2 | 75.8 | **77.2** | 65.9 | **82.5** |
| S-MMSE | **89.6** | **90.9** | 83.3 | **94.9** | **78.1** | 67.5 | **74.0** | 82.3 |

**Table 4**
Results for AD *vs*. MCI *vs*. NC classification with models trained on ADNI-1 and tested on ADNI-2.

| Method(%) | AD *vs*. MCI *vs*. NC | | | |
|---|---|---|---|---|
| | ACC | $ACC_{AD}$ | $ACC_{MCI}$ | $ACC_{NC}$ |
| C3D | 53.6 | 54.3 | 25.1 | **75.5** |
| G-CAT | 54.8 | **72.8** | 23.0 | 62.5 |
| S-MMSE | **56.6** | 59.8 | **41.3** | 65.8 |

**Table 6**
A brief description of the studies using ADNI-1 as training dataset and evaluate on the ADNI-2 for AD *vs*. NC classification.

| Reference(%) | Method | ACC | SEN | SPE | AUC |
|---|---|---|---|---|---|
| VBM [38] | VBM+SVM | 80.5 | 77.4 | 83.0 | 87.6 |
| ROI [39] | Tissue+SVM | 79.2 | 78.6 | 79.6 | 86.7 |
| LBM [27] | Landmark+SVM | 82.2 | 77.4 | 86.1 | 88.1 |
| wH-FCN [30] | Hierarchical CNN | **90.3** | 82.4 | 86.5 | **95.1** |
| G-CAT | Group Comaparative | 89.3 | 89.5 | **88.7** | 94.2 |
| S-MMSE | MMSE Ranking | **89.6** | 90.9 | 83.3 | 94.9 |

**Table 7**
Comparison of S-MMSE's effects on AD *vs*. NC task under different thresholds.

| Threshold | ACC | SEN | SPE | AUC |
|---|---|---|---|---|
| 0 | 90.9 | 93.1 | 92.6 | 96.3 |
| 1 | 90.1 | 93.7 | 91.6 | 96.5 |
| 2 | 92.4 | **97.7** | 91.4 | 96.8 |
| 3 | 91.3 | 92.0 | **94.7** | 97.2 |
| 4 | **93.5** | 96.6 | 93.4 | **97.8** |
| 5 | 85.9 | 84.6 | 93.7 | 92.4 |

the discriminative information extracting on AD *vs*. NC. We further noticed that the G-CAT performed better than the MMSE-based model on the MCI *vs*. NC task, which shows that the large overlap interval of the scores of MCI and NC could lead to deviations in the ranking information. In this case, the G-CAT based module is more effective for differential information extraction.

In order to further verify the effect of the model, we conducted a three category classification experiment of AD *vs*. MCI *vs*. NC. It can be seen accurately distinguishing three classifications is challenging, which is more difficult than MCI vs. NC. From Table 4, we can see S-MMSE yield the better result in the term of AC. In addition, it has a better recognition effect on MCI recognition when compared with the other methods. We can further see G-CAT performs best in identifying AD among the three comparison methods. In general, both G-CAT and S-MMSE are better than C3D on

three category classification task, which further illustrates the effectiveness of our proposed module.

### 3.2. Comparison with other methods in literature

We compared our proposed method with three conventional methods including VBM based methods, ROI (region of interest) based methods and several deep learning methods for early AD diagnosis. These models were also trained on the ADNI-1 dataset and evaluate on MIRIAD or ADNI-2 for AD *vs*. NC classification.

1. VBM based methods [38]. In the VBM methods, using the professional software processing to obtain the local GM tissue den-

**Table 5**
A brief description of the studies using ADNI-1 as training dataset and evaluate on the MIRIAD for AD *vs*. NC classification.

| Reference(%) | Method | ACC | SEN | SPE | AUC |
|---|---|---|---|---|---|
| VBM [38] | VBM+SVM | 88.4 | 91.3 | 82.6 | 92.1 |
| ROI [39] | Tissue+SVM | 87.0 | 91.3 | 82.6 | 91.8 |
| DSML-1[40] | Multi-Channel | 91.6 | 95.4 | 83.7 | 93.2 |
| DSML [40] | Multi-Channel+Demographic | 91.8 | **96.5** | 85.4 | 95.9 |
| $DML^2$-1[40] | Multi-Task+Multi-Channel | 92.0 | 96.3 | 89.8 | 96.9 |
| PDFC [41] | Patch+CNN | 91.3 | 89.1 | **95.6** | 97.1 |
| LDNFL [41] | Landmark+Direct Acyclic Network | 92.7 | 91.3 | **95.6** | **97.6** |
| G-CAT | Group Comaparative | 92.8 | 94.3 | 94.8 | 97.1 |
| S-MMSE | MMSE Ranking | **93.5** | **96.6** | 94.9 | **97.8** |

sity of brains' sMRI. Based on these features, a support vector machine (SVM) was constructed for the classification task.

2. ROI based methods [39]. The sMRI could be segmented into three types, GM, WM, and CSF. ROIs are extracted from the GM tissue as the feature for the subjects' sMRI. Similarly, an SVM model with a linear kernel was used for classification.

3. DSML [40]. DSML is a deep single-task multi-channel learning model using demographic information and DSML-1 is the single task model without demographic. $DML^2$-1 [40]. $DML^2$-1 is a deep multi-task multi-channel learning without using demographic information.

4. LBM: This method based on patch-level feature extraction. All the patches extracted from a pre-defined landmark was represented an embedding vector to perform SVM based classification[27].

5. wH-FCN: A hierarchical network was construct by automatically identifying multi-scale discriminative locations for AD diagnosis [30].

6. PDFC: This method learns the representation of patches by trainging multiple CNN. The features of each patch are concatenated into a SVM with linear kernel. [41]

7. LDNFL: LDNFL is a direct acyclic graph (DAG) network based on anatomical landmarks [42] for the diagnosis of AD. The DAG network is effective in the representation of MRI [41].

The comparison results on MIRIAD were shown in Table 5. It is noting our proposed S-MMSE and C-CAT outperforms the traditional ways like VBM and ROI. When compaered with DSML and its variants, S-MMSE achieves the same performance in SEN. The accuracy rates of G-CAT and S-MMSE surpass the remaining methods. In addtion, S-MMSE achieves an AUC of 97.8 % and G-CAT achieve an AUC of 97.1 %, which are comparable with the method based on landmark, like PDFC. The results on ADNI-2 were shown in Table 6. Comparison results indicating that our proposed model has a certain improvement compared with the traditional methods. The performance of G-CAT and S-MMSE on ADNI2 is still very similar, which is reflected in ACC, SEN and AUC. We found that our proposed S-MMSE achieves 94.9% in AUC. Although this is not the best, it is also very close to wh-FCN based on the landmarks with prior knowledge. In addition, our S-MMSE has the highest SEN compared to other methods.

### 3.3. Ablation experiment

#### 3.3.1. Effect of different scores

For S-MMSE model, we need to set the threshold $\xi$ empirically, thus we explored the effects of proposed models under different thresholds. Since the MMSE of AD and NC subjects are distributed in two distant intervals and their volatility is also small, we set the threshold from 0 to 5, and the experimental results were shown in Table 7. It can be seen from the experimental results that when the threshold is set to 4, the S-MMSE model performs best on the AD *vs.* NC task. This further demonstrates that the rank information between MMSE is beneficial to improve the performance of the model, especially when the difference between subjects' MMSE are large. According to the AUC change diagram shown in the Table 7, we found that the model performs better on the AD *vs.* NC task, when the threshold is set around 2–4.

#### 3.3.2. Effect of different image size

The second ablation experiment is to test the effects of different input image sizes used in our study. Considering the different sizes of each MRI could affect the model's results, we compared the model results at different resolutions and reported them in Table 8. In our experiment, we respectively set the size of the picture as $90 \times 90 \times 90$, $120 \times 120 \times 120$ and $150 \times 150 \times 150$ for the classification between AD and NC. The comparison demonstrates the

classification result of S-MMSE and G-CAT were improved slightly when the image size was set $90 \times 90 \times 90$. In general, when we set the image size $120 \times 120 \times 120$ or $150 \times 150 \times 150$, the model has the similar effects as the previous result, but requires more GPU memory.

#### 3.3.3. Evaluate the impact of hyper-parameters

We tested the impact of our hyper-parameters $\alpha$ and $\beta$ of G-CAT and S-MMSE under different weights on MIRIAD and ADNI-2 respectively. It can be seen from the Fig. 10 that when the hyper-parameters begin to increase, the performance of the model begins to decrease to varying degrees. It mainly because the cross-entropy loss used for classification is the main part and should account for a large proportion, while the loss function for contrast learning should account for a small part of the weight in the total loss function. We find that when the weight $\alpha, \beta$ is set to 0.2, the model performs best.

## 4. Discussion

### 4.1. Discussion on ambiguous samples

We analyzed the samples that were predicted incorrectly by the baseline model, those samples are usually of one's interest in clinical diagnosis. We found that the MMSE of those samples is very likely to distribute between 21–27. In order to better illustrate the effectiveness of our proposed model, we further visualize the MMSE distribution of the samples that were predicted correctly by the proposed model, and wrongly predicted by C3D. In the following, we will describe the predictive advantage of our proposed models including G-CAT and S-MMSE in detail, compared with the C3D model. For the AD *vs.* NC task, we integrate the results of the model on MIRIAD and the results on ADNI-2 and display them together in Figs. 9 (a) and (b).

#### 4.1.1. Discussion on the results of G-CAT

For G-CAT based methods on AD *vs.* NC (Fig. 9 (a)), one can find that the blue scatters are rare, which shows the baseline model and our proposed model both have good predictive ability for CN. There is a lot of red points in Fig. 9 (c) and they are roughly distributed in the 21–25, which shows the baseline model has the poor predictive ability for AD subjects with scores between 21 and 25, while our model can make more accurate predictions for subjects with scores in this interval. For MCI *vs.* NC, our proposed model accurately identified MCI subjects concentrated around 28 and NC subjects concentrated around 29–30. The above results demonstrate that G-CAT module provides more discriminative features by comparing the differences between classes.

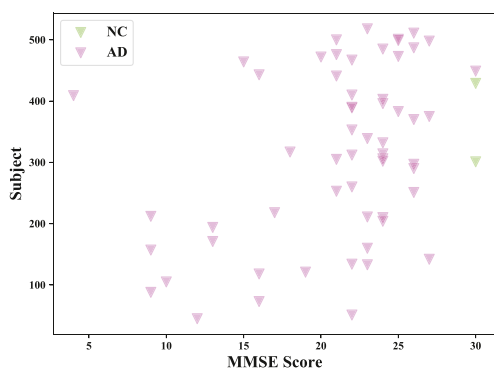#### 4.1.2. Discussion on the results of S-MMSE

Similar to the G-CAT, as can be seen from the Fig. 9 (b), S-MMSE also performs well on AD *vs.* NC and has a strong recognition ability for AD patients. We can further conclude that due to the large difference in the distribution of MMSE between AD and NC, the difference between the scores has covered the information of the category, and the effect of using the ranking information between the MMSE can be roughly equivalent for G-CAT. The MMSE of MCI subjects fluctuate greatly, and there is a large overlap with NC subjects. It can be seen from the Fig. 9 (d) that our model can accurately predict MCIs distributed in 27–29, which demonstrates that S-MMSE has a strong ability to predict confused samples.

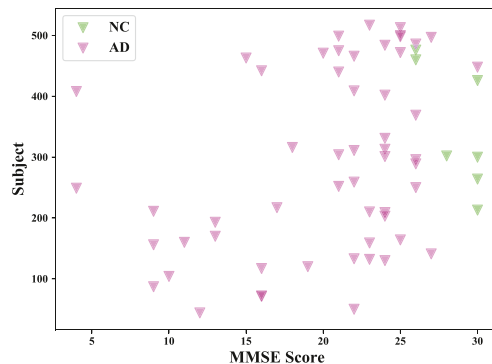### 4.2. Discussion on the visualization of representations

Another important classification evaluation method is to generate a visual representation of the sample in a two dimensional

**Table 8**
Results of classification of AD *vs*. NC on MIRIAD with different input image size. Learning models are trained on ADNI-1.
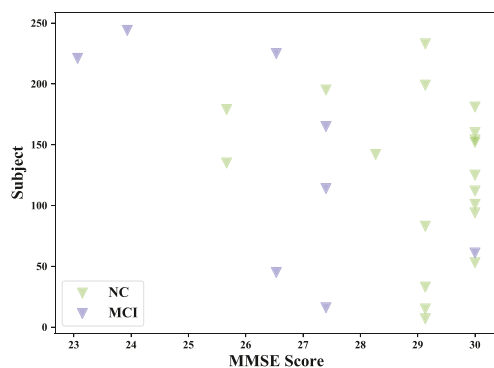
| Method(%) | G-CAT | | | | S-MMSE | | | |
|---|---|---|---|---|---|---|---|---|
| | ACC | SEN | SPE | AUC | ACC | SEN | SPE | AUC |
| 90 × 90 × 90 | 92.8 | 94.3 | 94.8 | 97.1 | 93.5 | 97.6 | 94.9 | 97.8 |
| 120 × 120 × 120 | 91.6 | 92.5 | 94.7 | 96.6 | 93.5 | 94.3 | 95.9 | 97.3 |
| 150 × 150 × 150 | 92.0 | 93.7 | 94.2 | 96.7 | 92.4 | 94.8 | 93.8 | 96.8 |



(a) G-CAT on AD *vs*. NC classification



(b) S-MMSE based on AD *vs*. NC classification



(c) G-CAT on MCI *vs*. NC classification



(d) S-MMSE based on MCI *vs*. NC classification

**Fig. 9.** The distribution of the specific samples which are predicted correctly by our proposed model and are predicted wrongly by C3D on both AD *vs*. NC and MCI *vs*. NC classification tasks.

space. The feature representations of the samples were obtained by the output of the first fully connected layer. Based on the visualization tool t-SNE [43], each subject with high dimension features is mapped as a two dimensional vector. As shown in Fig. 11, for subjects are labeled as different categories, we use different colors on the corresponded points. Therefore a good visualization result is that the points of the same color gather more closely [44]. Fig. 11 (a)–(c) show the visualizations of samples in MIRIAD on AD *vs*. NC classifications. It can see that the results of C3D are not satisfactory because the points belonging to different categories are mixed. For G-CAT and S-MMSE, points with different categories are relatively separated by a certain distance, and points with the same category are clustered together. From Fig. 11 (d)–(f), it can be also seen that G-CAT and S-MMSE have better visualization results than C3D for AD *vs*. NC classification on ADNI-2. In addition, the S-MMSE performs best among the three methods which further verified that the MMSE ranking information is beneficial for the AD *vs*. NC classification.

### 4.3. Literature evidence

On the AD *vs*. NC task, we found that S-MMSE achieved better performance by setting the threshold between 2 to 4. Several studies suppose that a decline of 3 points in MMSE infers morphological changes in brain sMRI, which is associated with dementia [45–47]. Thus, our findings are consistent with this hypothesis, which could be a useful clinical basis for the early diagnosis of AD. It is also worth mentioning that MMSE score is only used during training process, only sMRI is required during prediction. i.e. The proposed method does not require MMSE score as input for clinical usage.

### 4.4. Limitations and future work

While our proposed model performs well on three datasets, several limitations should be addressed in the future to further improve its performance. First, our current model was implemented to mine subtle differences between samples, but unable to locate
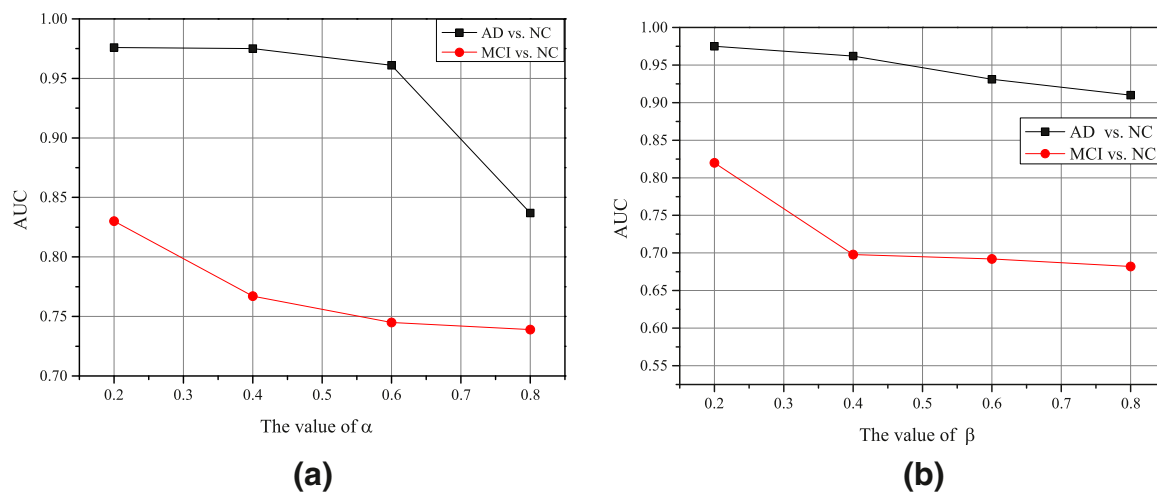
**Fig. 10.** Results of effects under different (a) $\alpha$ and (b) $\beta$. AD *vs.* NC is the classification on MIRIAD and MCI *vs.* NC is the classification on ADNI-2.
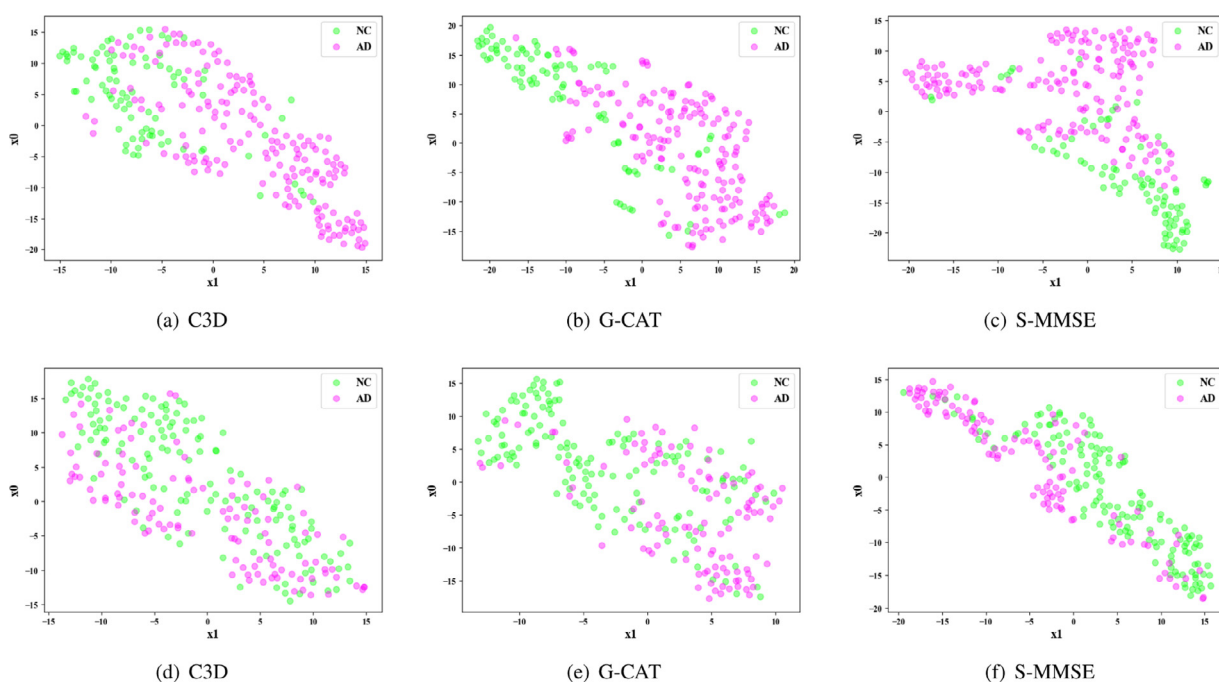


**Fig. 11.** Visualization of subjects in MIRIAD and ADNI-2. Each point indicates one subject. The color of a point indicates the category of the subject including AD and NC. $x0$ and $x1$ are the representation of the embedding vector after dimensionality reduction. (a)–(c): Visualizations of subjects in MIRIAD on AD *vs.* NC classification. (d)–(f): Visualizations of subjects in ADNI-2 on AD *vs.* NC classification.

and visualize these features. It is an interesting and promising direction to materialize and visualize these extracted features. Secondly, we do not take into account the changes of a single individual over time but uses the state of the subject at a specific time node. We believe that the timing information of the subject can be more conducive to the clinical diagnosis of AD. Finally, datasets under different scan types (i.e., 1.5T scanners and 3T scanners) are also directly used in this research. We also believe the domain adaptation model [48] could further its generalization capability. In future work, we consider further visualizing the features we extracted using class activation mapping [49]. On the basis of our model, consider the characteristics of a single subject change information over time to improve the recognition ability of the model.

**Declaration of Competing Interest**

All authors declared no conflict of interests.

**CRediT authorship contribution statement**

**Hezhe Qiao:** Conceptualization, Methodology, Software, Writing – original draft. **Lin Chen:** Conceptualization, Methodology, Writing – review & editing. **Zi Ye:** Data curation, Writing – original draft. **Fan Zhu:** Formal analysis, Supervision.

**Acknowledgment**

in analysis or writing of this report. The MIRIAD dataset is made available through the support of the UK Alzheimer's Society (Grant RF116). The original data collection was funded through an unrestricted educational grant from GlaxoSmithKline (Grant 6GKC).

## References

[1] A. Burns, Alzheimer's disease: on the verges of treatment and prevention, Lancet Neurol 8 (1) (2009) 4–5.

[2] M. Tanveer, B. Richhariya, R.U. Khan, et al., Machine learning techniques for the diagnosis of alzheimer's disease: a review, ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 16 (1) (2020) 1–35.

[3] A. Abeysinghe, R. Deshapriya, C. Udawatte, Alzheimer'S disease; a review of the pathophysiological basis and therapeutic interventions, Life Sci. (2020) 117996.

[4] N. Bregman, G. Kavé, E. Zeltzer, et al., Memory impairment and alzheimer's disease pathology in individuals with MCI who underestimate or overestimate their decline, Int J Geriatr Psychiatry 35 (5) (2020) 581–588.

[5] X. Ji, et al., Brainstem atrophy in the early stage of alzheimer's disease: a voxel-based morphometry study, Brain Imaging Behav 15 (1) (2021) 49–59.

[6] S.A. Mofrad, et al., Cognitive and MRI trajectories for prediction of alzheimer's disease, Sci Rep 11 (1) (2021) 1–10.

[7] R. Cui, M. Liu, G. Li, Longitudinal analysis for Alzheimer's disease diagnosis using RNN, in: Proceedings of the IEEE 15th International Symposium on Biomedical Imaging (ISBI), 2018, pp. 1398–1401.

[8] C.C. Luk, et al., Alzheimer'S disease: 3-dimensional MRI texture for prediction of conversion from mild cognitive impairment, Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 10 (1) (2018) 755–763.

[9] H. Fuse, et al., Detection of Alzheimer's disease with shape analysis of MRI images, in: Proceedings of Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS), 2018, pp. 1031–1034.

[10] A.L. Benedet, et al., SNAP25 Reflects amyloid-and tau-related synaptic damage: associations between PET, VBM and cerebrospinal fluid biomarkers of synaptic disfunction in the alzheimer's disease spectrum: neuroimaging: imaging the human synapse in AD, Alzheimer's & Dementia 16 (2020) e046358.

[11] S. Leandrou, D. Lamnisos, I. Mamais, et al., Assessment of alzheimer's disease based on texture analysis of the entorhinal cortex, Front Aging Neurosci 12 (2020) 176.

[12] T. Guo, D. Korman, S.L. Baker, et al., Longitudinal cognitive and biomarker measurements support a unidirectional pathway in alzheimer's disease pathophysiology, Biol. Psychiatry 89 (8) (2021) 786–794.

[13] R. Jain, N. Jain, A. Aggarwal, et al., Convolutional neural network based alzheimer's disease classification from magnetic resonance brain images, Cogn Syst Res 57 (2019) 147–159.

[14] F. Previtali, P. Bertolazzi, G. Felici, et al., A novel method and software for automatically classifying alzheimer's disease patients by magnetic resonance imaging analysis, Comput Methods Programs Biomed 143 (2017) 89–95.

[15] X. Zhang, Y. Yang, T. Li, et al., CMC: A consensus multi-view clustering model for predicting alzheimer's disease progression, Comput Methods Programs Biomed 199 (2021) 105895.

[16] P. Cao, X. Liu, J. Yang, et al., Nonlinearity-aware based dimensionality reduction and over-sampling for AD/MCI classification from MRI measures, Comput. Biol. Med. 91 (2017) 21–37.

[17] M.A. Ebrahimighahnavieh, S. Luo, R. Chiong, Deep learning to detect alzheimer's disease from neuroimaging: a systematic literature review, Comput Methods Programs Biomed 187 (2020) 105242.

[18] S. Liang, Y. Gu, Computer-aided diagnosis of alzheimer's disease through weak supervision deep learning framework with attention mechanism, Sensors 21 (1) (2021) 220.

[19] Q. Li, Y. Zhang, H. Liang, et al., Deep learning based neuronal soma detection and counting for alzheimer's disease analysis, Comput Methods Programs Biomed 203 (2021) 106023.

[20] M. Liu, F. Li, H. Yan, et al., A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in alzheimer's disease, Neuroimage 208 (2020) 116459.

[21] C. Lian, et al., Attention-guided hybrid network for dementia diagnosis with structural MR images, IEEE Trans Cybern (2020).

[22] J. Liu, M. Li, Y. Luo, et al., Alzheimer's disease detection using depthwise separable convolutional neural networks, Comput Methods Programs Biomed 203 (2021) 106032.

[23] J. Wen, E. Thibeau-Sutre, M. Diaz-Melo, et al., Convolutional neural networks for classification of alzheimer's disease: overview and reproducible evaluation, Med Image Anal 63 (2020) 101694.

[24] D. Jin, J. Xu, K. Zhao, et al., Attention-based 3D convolutional network for Alzheimer's disease diagnosis and biomarkers exploration, in: Proceedings of the IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019, pp. 1047–1051.

[25] R. Cui, M. Liu, Hippocampus analysis by combination of 3-d densenet and shapes for alzheimer's disease diagnosis, IEEE J Biomed Health Inform 23 (5) (2018) 2099–2107.

[26] S. Korolev, A. Safiullin, M. Belyaev, et al., Residual and plain convolutional neural networks for 3D brain MRI classification, in: Proceedings of the IEEE 14th International Symposium on Biomedical Imaging (ISBI), 2017, pp. 835–838.

[27] J. Zhang, Y. Gao, Y. Gao, et al., Detecting anatomical landmarks for fast alzheimer's disease diagnosis, IEEE Trans Med Imaging 35 (12) (2016) 2524–2533.

[28] L. Wang, Y. Liu, X. Zeng, et al., Region-of-interest based sparse feature learning method for alzheimer's disease identification, Comput Methods Programs Biomed 187 (2020) 105290.

[29] M. Liu, J. Zhang, E. Adeli, et al., Landmark-based deep multi-instance learning for brain disease diagnosis, Med Image Anal 43 (2018) 157–168.

[30] C. Lian, M. Liu, J. Zhang, et al., Hierarchical fully convolutional network for joint atrophy localization and alzheimer's disease diagnosis using structural MRI, IEEE Trans Pattern Anal Mach Intell 42 (4) (2020) 880–893.

[31] M. Liu, J. Zhang, C. Lian, et al., Weakly supervised deep learning for brain disease prognosis using MRI and incomplete clinical scores, IEEE Trans Cybern 50 (7) (2020) 3381–3392.

[32] K. Aderghal, A. Khvostikov, A. Krylov, et al., Classification of Alzheimer disease on imaging modalities with deep CNNs using cross-modal transfer learning, in: 2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS), 2018, pp. 345–350.

[33] T.D. Vu, N.H. Ho, H.J. Yang, et al., Non-white matter tissue extraction and deep convolutional neural network for alzheimer's disease detection, Soft comput 22 (20) (2018) 6825–6833.

[34] C. Song, Y. Huang, W. Ouyang, et al., Mask-guided contrastive attention model for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 1179–1188.

[35] C.R. Jack Jr, M.A. Bernstein, N.C. Fox, et al., The alzheimer's disease neuroimaging initiative (ADNI): MRI methods, Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine 27 (4) (2008) 685–691.

[36] I.B. Malone, D. Cash, G.R. Ridgway, et al., MIRIAD Public release of a multiple time point alzheimer's MR imaging dataset, Neuroimage 70 (2013) 33–36.

[37] K. Hara, H. Kataoka, Y. Satoh, Learning spatio-temporal features with 3D residual networks for action recognition, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 3154–3160.

[38] J. Ashburner, K.J. Friston, Voxel-based morphometry the methods, Neuroimage 11 (6) (2000) 805–821.

[39] D. Zhang, Y. Wang, L. Zhou, Multimodal classification of alzheimer's disease and mild cognitive impairment, Neuroimage 55 (3) (2011) 856–867.

[40] M. Liu, J. Zhang, E. Adeli, et al., Joint classification and regression via deep multi-task multi-channel learning for alzheimer's disease diagnosis, IEEE Trans. Biomed. Eng. 66 (5) (2018) 1195–1206.

[41] T. Zhu, C. Cao, Z. Wang, G. Xu, J. Qiao, Anatomical landmarks and DAG network learning for alzheimer's disease diagnosis, IEEE Access 8 (2020) 206063–206073.

[42] M. Liu, J. Zhang, D. Nie, P.T. Yap, D. Shen, Anatomical landmark based deep feature representation for MR images in brain disease diagnosis, IEEE J Biomed Health Inform 22 (5) (2018) 1476–1485.

[43] L. Van der Maaten, G. Hinton, Visualizing data using t-SNE, Journal of Machine Learning Research 9 (2008) 2579–2605.

[44] D. Wang, P. Cui, W. Zhu, Structural Deep Network Embedding, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1225–1234.

[45] A. Hensel, M.C. Angermeyer, S.G. Riedel-Heller, Measuring cognitive change in older adults: reliable change indices for the mini-mental state examination, Journal of Neurology, Neurosurgery Psychiatry 78 (12) (2007) 1298–1303.

[46] P.K. Crane, A. Carle, L.E. Gibbons, et al., Development and assessment of a composite score for memory in the alzheimers disease neuroimaging initiative (ADNI), Brain Imaging Behav 6 (4) (2012) 502–516.

[47] G.I. Allen, N. Amoroso, C. Anghel, et al., Crowdsourced estimation of cognitive decline and resilience in alzheimer's disease, Alzheimer's Dementia 12 (6) (2016) 645–653.

[48] A. Rozantsev, M. Salzmann, P. Fua, Beyond sharing weights for deep domain adaptation, IEEE Trans Pattern Anal Mach Intell 41 (4) (2018) 801–814.

[49] B. Zhou, A. Khosla, A. Lapedriza, et al., Learning deep features for discriminative localization, in: Proceedings of the IEEE Conference on Computer Vision and Pattern recognition (CVPR), 2016, pp. 2921–2929.