

AMIM: An Adaptive Weighted Multimodal Integration Model for Alzheimer's Disease Classification

Dewen Ding¹, Xianhua Zeng², Xinyu Wang³ and Jian Zhang⁴
Chongqing University of Posts and Telecommunications
Chongqing Key Laboratory of Image Cognition, China, Chongqing

Abstract—Alzheimer's disease (AD) is an irreversible neurological disorder, so early medical diagnosis is extremely important. Magnetic resonance imaging (MRI) is one of the main medical imaging methods used clinically to detect and diagnose AD. However, most existing computer-aided diagnostic methods only use MRI slices for model architecture design. They ignore informational differences between all slices. In addition, physicians often use multimodal data, such as medical images and clinical information, to diagnose patients. The approach helps physicians to make more accurate judgments. Therefore, we propose an adaptive weighted multimodal integration model (AMIM) for AD classification. The model uses global information images, maximum information slices and clinical information as data inputs for the first time. It adopts adaptive weights integration method for classification. Experimental results show that our model achieves an accuracy of 99.00% for AD versus normal controls (NC), and 82.86% for mild cognitive impairment (MCI) versus NC. The proposed model achieves best classification performance in terms of accuracy, compared with most state-of-the-art methods.

Keywords—MRI; global information images; maximum information slices; adaptive weights; integration method

I. INTRODUCTION

Alzheimer's disease, a chronic neurodegenerative disease causing the death of nerve cells and tissue loss throughout the brain, usually starts slowly and worsens over time. AD is expected to affect 1 out of 85 people in the world by the year 2050 [1]. The progression of AD gradually leads to memory degradation and impairment of cognitive function, eventually leading to irreversible neuronal damage [2]. Although no treatment has been proven to be effective in preventing the progression of AD [3], the early diagnosis of AD remains important for subsequent treatment to delay the onset of cognitive symptoms [4].

Since 2013, deep learning has begun to gain considerable attention in AD detection research, with the number of published papers in this area increasing drastically since 2017 [5]. Early unsupervised methods used autoencoders or restricted Boltzmann machine methods to extract features that were then used for the classification of Alzheimer's disease [6]–[8]. Supervised learning applied to the diagnosis of Alzheimer's disease has been particularly well studied compared to unsupervised methods. Convolutional neural networks (CNNs) are the most successful deep models for image analysis, aiming to make better use of spatial information by taking 2D/3D images as input and extracting features by stacking several

convolutional layers; the result is a hierarchy of progressively abstracted features [9], [10]. Most studies on Alzheimer's disease have been mainly architected by 2D CNNs or 3D CNNs as depth models. A large number of studies have performed feature extraction of MRI slices by 2D CNNs for ADNI classification [11]–[16]. Since MRI provides 3D images, how to select MRI slices is a question worth thinking about. Meanwhile, 2D slices cannot contain all the information of 3D MRI, so it is missing the global information. 3D CNN is widely used for diagnosis of 3D MRI, which does not require slice selection and also contains global information. However, AD detection must take the whole image or some ROI as input [17], [18]. This results in a steep increase in the number of parameters, which can create problems such as large amount of computation, time-consuming, and overlapping data. The joint 2D-3D CNN [19]–[21] first performs 3D feature extraction by multiple 3D data inputs, and then obtains the final classification result by 2D CNN. Likewise, there are also problems such as large amount of computational cost and time consuming. In addition, in the early detection of Alzheimer's disease, the degree of brain atrophy is less variable, and assessment by a single modality of MRI alone may have a certain bias; a combined assessment with multiple modalities will yield a relatively more accurate diagnosis.

Based on slice-level classification, there is a lack of three-dimensional spatial information and the subjective uncertainty of slice selection. We proposed to effectively superimpose all slices to generate a dynamic 2D picture containing multiple slice information changes, i.e. a global information image. At the same time, the slice with the largest amount of information is also selected through the method of image entropy, and the clinical information is used as the input of the multi-dimensional feature auxiliary integrated approach. From our experiment, the proposed AMIM model has improved the performance significantly. The main contributions of our study are summarized in the following three folds.

- We propose an adaptive weighted multimodal ensemble model. The model uses an adaptive weighted method to optimize the different branches weights. It can effectively reduce the large amount of computational cost and time, compared with the grid search method.
- For the first time, we propose a new MRI image preprocessing method, which uses dynamic images and maximum information entropy slices as the input

of MRI images; at the same time, clinical information modality is introduced to obtain better classification performance.

- A comprehensive evaluation is conducted on ADNI dataset. Experiments show that our method achieves best classification performance in terms of accuracy, compared with most state-of-the-art methods.

The rest of the paper is structured in sections and represented as follows. In Section II, related work describes the research status of Alzheimer’s disease classification in detail. Section III introduces the structure and algorithm of the AMIM model. Section IV introduces the classification performance of the model on the ADNI dataset. Section V discusses the performance analysis of different views. Section VI concludes the paper.

II. RELATED WORK

With the rapid development of deep learning since 2012, there are more and more researches on the diagnosis of Alzheimer’s disease. Researches based on Alzheimer’s disease classification can be divided into the following branches according to input: ROI level, Patch level, 3D Subject level and 2D slice level. With ROI models [22], [23], manual selection of regions is required to extract the region of interest of the original brain image as the input of CNN model, which is a time-consuming task. With patch models [19]–[21], multiple patches can be obtained from the entire 3D MRI, but there is a problem of data overlap. It is much more straightforward and desirable to use the entire image as input. At the 3D subject level, Korolev et al. [17] adopted 3D VGG and 3D ResNet as the backbone network for feature extraction, but the classification accuracy was only more than 80. Spasov et al. [18] proposed a method combining 3D MRI with clinical information, which can obtain good classification results. However, regardless of single mode or multi-mode 3D MRI, there is a large amount of calculation and long running time. In 2D slice classification method, it can reduce the number of hyperparameters to a certain extent. Due to the small sample size of medical dataset, Hon et al. [11] proposed to apply transfer learning to the classification of Alzheimer’s disease. 32 slices were taken from each object as the dataset, and the model performed well. However, this result was only for the image level, without considering the subject level. Islam et al. [12] proposed a deep convolutional neural network for the diagnosis of Alzheimer’s disease using brain MRI data analysis, and obtained good classification results. Zhang and colleagues [13] performed a systematic evaluation of CNN models with different structures and capacities, and the experimental results showed that the advanced structural models with medium capacity performed better than the models with maximum capacity. Good results have also been obtained. However, these methods are based on 2D images and cannot contain all the information of brain scan. They ignore the spatial information of 3D. The same situation exists in other slice classification studies [14]–[16]. According to the above analysis, we propose a new AD classification network architecture, AMIM, combining 2D-3D MRI and clinical information.

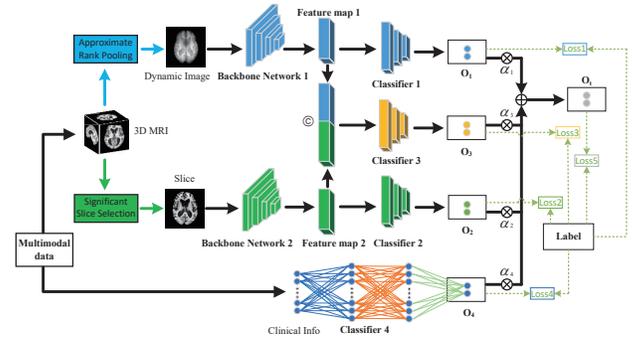


Fig. 1. The AMIM for alzheimer’s disease classification based on MRI and clinical information

III. METHODS

In this paper, we propose an AMIM model that combines 2D-3D MRI to solve the problem of missing 3D spatial information for slice based classification. Clinical information is also introduced as the input of another modality. The model uses an adaptive weighted method to learn the weight shares of different classifier. Its architecture is shown in Fig. 1. Our proposed method is flexible and can in principle integrate other imaging modalities, such as positron emission computed tomography (PET), as well as other different clinical datasets. With the inspiration of the idea of transfer learning, we use the classical neural network pre-trained by ImageNet and removing the last classification layer as the backbone network of feature extraction [24]. Resnet18 as a backbone network will be introduced here.

In the following, we present our method in four parts. First, we introduce the dynamic image generation method and the maximum information entropy slicing method, respectively. Then, we present the adaptive weighted multimodal integration method. Finally, we introduce the training and optimization.

A. Dynamic Image

In the non-medical field, a popular method to represent a series of images is to apply a temporal pooling operator to the features extracted at individual images, for instance, temporal templates [25], ranking functions [26] and other traditional pooling operator [27]. We use the Z-dimension of the 3D MRI as the temporal dimension of the video to extract a fixed slice representation of each object. Since the extracted fixed representation retains all the dynamic characteristics of slices (i.e. the changes from slice to slice), we call it dynamic image. We calculate the coefficient θ_t of slice I_t and assume that the feature vector of this slice is V_t . Multiply this coefficient with the average of all feature vectors from V_1 to V_t to get the new feature vector and finally accumulate the new feature vectors to get the final dynamic image. See Fig. 2 for an example. The calculation formula is as follows.

$$I_D = \sum_{t=1}^T \theta_t \psi(V_t) \quad (1)$$

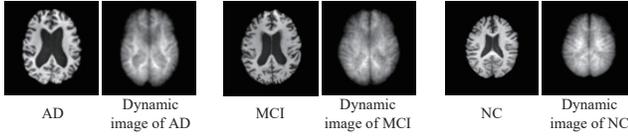


Fig. 2. MRI sample slices and corresponding dynamic images of AD, MCI and NC participants

where $\psi(V_t) = \frac{1}{t} \sum_{\tau=1}^t V_\tau$ is the slices average of features up to slice t , and $\theta_t = 2t - T - 1$ is the coefficient associated to feature $\psi(V_t)$.

The algorithm for processing dynamic images is shown in Algorithm 1.

Algorithm 1 Algorithm for Obtaining Dynamic Images

Input: Origin 3D MRI: $\{X_{MRI}^1, X_{MRI}^2, \dots, X_{MRI}^N\}$.

Output: Dynamic images $\{I_D^1, I_D^2, \dots, I_D^N\}$.

- 1: **for** $i = 1$ to N **do**
 - 2: $I_O^j \leftarrow \text{slice}(X_{MRI}^i), j = \{1, 2, \dots, T\}$ // Obtaining original axial slices.
 - 3: $V_j \leftarrow \text{quantify}(I_O^j)$ // Obtaining feature vectors.
 - 4: **for** $t = 1$ to T **do**
 - 5: $\theta_t \leftarrow 2t - T - 1$ // Calculating the coefficient θ_t of slice I_O^t .
 - 6: $\psi(V_t) \leftarrow \frac{1}{t} \sum_{\tau=1}^t V_\tau$ // Calculating the vector mean $\psi(V_t)$ of the features vectors V_1 to V_t .
 - 7: **end for**
 - 8: $I_D^i \leftarrow \sum_{t=1}^T \theta_t \psi(V_t)$ // All new feature vectors are summed to get the final dynamic image by Eq.(1).
 - 9: **end for**
-

B. Slice with Maximum Information

Typically, there are a large number of slices to choose from in 3D MRI scan. One method of slices selection is to manually select slices based on the highest similarity of anatomical features without knowing the clinical diagnosis information [28]. However, this approach needs to be chosen by professionals, which will cost a lot of labor and be subjective. Instead, we use image entropy to extract the most informative slices to train the network. Therefore, we will calculate the image entropy of each slice. Generally speaking, for a set of M symbols with probabilities P_1, P_2, \dots, P_m . Entropy can be calculated as [29] :

$$H = - \sum_{i=0}^{255} P_i \log P_i \quad (2)$$

where H is the one-dimensional gray-scale entropy and P_i is the proportion of grayscale value i .

C. Adaptive Weighted Multimodal Integration

In this section, we introduce the composition of classifiers and the adaptive weighted integration of the classifier, respectively.

For these classifiers we use the same composition structure. We add relu activation function and dropout between each layer of mapping to reduce the potential overfitting risks. Specifically the feature images are first dropout, and then through three layers of mapping, relu and dropout are added between each layer of mapping. Finally, softmax activation function is used to obtain the category probability value. Dropout set to 0.5. The output expression of each of these classifiers is shown as follows:

$$O_i = \varphi^i(F_i, C_i) \quad i \in \{1, 2, 3\} \quad (3)$$

where F_i stands for the input of the feature map, C_i is the weight parameter of the i -th classifier, $\varphi^i(F_i, C_i)$ represents a function to be learned in an effort to transform the input, F_i , to probability value O_i .

$$O_4 = \varphi^4(X_{cli}, C_4) \quad (4)$$

where O_4 is the probability value in the 4-th classifier, X_{cli} is the clinical information. C_4 is the weight parameter of the 4-th classifier. φ^4 denotes the operation of the 4-th classifier. We normalized for clinical characteristics, i.e. demographic, neuropsychological, and the apolipoprotein E (APOE4) genotyping data. They all followed the same feature scaling procedure, with values normalized between [0, 1] for each independent clinical factor.

We propose an ensemble learning method with adaptive weights to improve the performance of the model and the confidence of prediction. For the probability values of multiple classifiers, we use the integration method of soft voting for the final output. Suppose we have M classifiers, the soft voting can be computed as:

$$O_t = \sum_{i=1}^M \alpha_i O_i \quad (5)$$

where O_i is the probability value in the i -th classifier, α_i represents the weight given to the i -th classifier. O_t stands for the total output after overall soft voting integration. First we initialize the weights. To automatically compute the hyperparameter α_i , we use a simple but effective approach: setting the hyperparameter α_i as a trainable parameter in order to automatically and adaptively coordinate the importance learning of each attribute task. When multiple branch tasks are learned simultaneously, the ‘‘important’’ branches should be given high weights (i.e. α_i) to increase the loss size of the corresponding branch. We take a small learning rate for updating the network parameters and automatically learn the score weights for different classifiers.

D. Training and Optimization

We use cross-entropy as the loss function. We construct loss function for the output of individual classifier. The loss function is:

$$L_i = - \frac{1}{N} \sum_{j=1}^N \left[y_j \cdot \log(O_i^j) + (1 - y_j) \cdot \log(1 - O_i^j) \right] \quad i \in \{1, 2, 3, 4\} \quad (6)$$

where the label $y_j = 0$ indicates that sample j is a negative sample and $y_j = 1$ indicates that sample j is a positive sample. N is the total number of samples in the data set. O_t^j denotes the output probability value of sample j of the i -th classifier. In the training phase, with the loss function constructed from the output of each classifier, i.e., Equation 6, we can optimize the network parameters in each of them.

For the backbone network, the weights are fixed in the first stage and are not optimized. The weights are unfrozen in the second stage, we first optimize the backbone network 1 and backbone network 2 under the loss functions constructed from the classifier 1 output and classifier 2 output, respectively. Then the backbone network is further optimized by the loss function 3.

After we get the output of each classifier, we can get the final output, which is Equation 5. Then the loss function of the integrated output is shown below:

$$L_5 = -\frac{1}{N} \sum_{j=1}^N \left[y_j \cdot \log \left(O_t^j \right) + (1 - y_j) \cdot \log \left(1 - O_t^j \right) \right] \quad (7)$$

where α_i is the weight of the output probability values of the i -th classifier. O_t^j represents the output probability value of the soft voting integration of sample j through multiple classifiers. The hyperparameter α_i is key part of the network. If we use the grid search method to obtain α_i , this will be time consuming. We use the adaptive weighted method to update the network weights. The loss combines these weights as an integrate one to supervise the process of network training by adopting the back-propagation algorithm.

IV. EXPERIMENTS

A. Dataset

We use the publicly available dataset from the Alzheimer's Disease Neuroimaging Initiative (ADNI) for our work. Specifically, we trained CNNs with the data from the "spatially normalized, masked, and N3-corrected T1 images" category. The brain MRI image size is $110 \times 110 \times 110$. Since a subject may have multiple MRI scans in the database, we use the most recent scan of each subject to avoid data leakage. All the data we used are summarized in Table I. Among them are 132 men and 92 women, aged between 55 and 90.3 years old. Friedman's ANOVA was used to test the difference in median age between groups, and the Fisher's exact test was used to test the gender interaction of the x group. These interactions are not statistically significant ($p > .05$). And the following data: demographic data (age, gender, education level), neuropsychological cognitive assessment tests such as the Dementia Rating Scale (CDRSB), Alzheimer's Disease Assessment Scale (ADAS11, ADAS13), Rey Auditory Verbal Learning Test (RAVLT), as well as APOE4 genotyping. All data used in this study is from baseline assessments.

B. Evaluation Metrics

The proposed AMIM method mainly validates the AD classification (AD vs. NC), MCI classification (MCI vs. NC). The performance was evaluated using three metrics. Namely, accuracy, the percentage of correctly predicted samples; F1,

TABLE I. DEMOGRAPHICS OF THE ADNI DATASET

	F/M	Education	Age	APOE4	MMSE	CDRSB	ADAS11	ADAS13	RAVLT IR	RAVLT PF
CN	28/28	16.45±2.95	73.11±5.64	0.16±0.23	28.59±1.82	0.44±1.09	6.32±4.23	9.85±6.19	41.68±10.46	47.48±32.84
MCI	48/75	15.80±2.67	72.95±7.39	0.34±0.34	26.29±3.34	2.23±1.98	11.25±6.45	17.64±9.37	31.50±11.58	66.98±33.84
AD	16/29	15.37±2.89	74.66±8.54	0.39±0.33	21.67±3.94	5.81±2.40	24.07±7.48	34.78±9.01	19.56±6.71	93.68±13.35

the harmonic mean of precision (Eq. 9) and recall (Eq. 10); Area Under Curve (receiver operating characteristic curve determined by true positive rate and false positive rate). These metrics are defined as:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (9)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (10)$$

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

where TP, TN, FP and FN stands for true positive, true negative, false positive and false negative, respectively. A higher value indicates better performance.

In the following, experiments are conducted to evaluate the performance of the proposed method. Specifically, Section IV-D focuses on testing the impact of unimodal and multimodal on the performance metrics of the experiments, respectively. Section IV-E1 aims to analyze the classification performance of the same dataset in different methods.

C. Complements

We use five-fold cross-validation for experiments. Since the proportion of data samples is unbalanced, a weighted loss function is used to ensure the balance of the samples. The loss function uses a cross-entropy loss function. Using the Adam optimizer, the learning rate is 10 to the negative 5-th power, except that the learning rate of L_5 is adjusted to 5×10^{-6} . The classifier performs linear mapping, it will perform a dropout of 0.5 to prevent overfitting. To activate relu, the last layer uses the softmax function to output the probability value of the category.

D. Single Modality vs. Multiple Modalities

In this section, the effects of unimodal and multimodal data on the model are presented separately. For MRI, resnet18, which has been pre-trained and removed the last classification layer, is used as the backbone network to extract features. The output of clinical information is obtained through a multi-layer perceptron network model. We performed single-modal experiments on dynamic images, slices, and clinical information. As shown in the Table II, we can see that in the column of AD versus NC, the evaluation index of clinical information is very high. Thus we made t-SNE visualization for the data, as shown in Fig. 3. The data distribution of AD group and NC group is shown on the left. It is found that the two types of data have obvious dividing lines. The distributions of the MCI and NC groups on the right do not have obvious dividing

TABLE II. COMPARISON OF PERFORMANCE OF SINGLE-MODAL AND MULTIMODAL CLASSIFICATION METHODS

Methods	AD vs. NC (%)			MCI vs. NC (%)		
	ACC	AUC	F1	ACC	AUC	F1
Dynamic	90.00	91.90	88.63	77.10	74.45	84.61
Slice	83.14	82.59	83.09	71.00	62.34	80.76
Clinical Info	97.00	98.03	97.41	77.67	78.65	85.53
Combined	92.00	93.62	91.06	74.29	72.39	82.26
Fusion	99.00	99.12	98.67	81.63	82.41	87.06

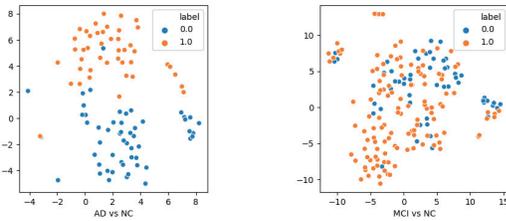


Fig. 3. Illustration of t-SNE result for cognitive scores. Blue dots: NC group. Orange dots: AD/MCI group

lines, and their evaluation indexes in Table II are not very satisfactory. This remains consistent with our results. Then we take the dynamic images and slices of medical imaging as input to obtain the integrated results of medical imaging. Specifically, the decision fusion of Output 1, Output 2 and Output 3 of the model in Fig. 1 is carried out. The specific experiment is introduced in the subsequent result chapter. It can be seen from the table that the integration results of slices combined with pictures of 3D spatial information changes are improved compared with single-mode medical imaging. Finally, we proposed the method, in which AD/NC accuracy reached 99.00%, MCI/NC accuracy also reached 81.63%.

E. Comparison of Different Methods

1) *Comparison With baseline methods:* We compared several other methods on the same dataset. Korolev et al. proposed a deep three-dimensional convolutional neural network structure for brain MRI scan classification [17]. In this work [18], structural magnetic resonance imaging (MRI), demography, neuropsychology and APOE4 gene were used as data inputs, and 3D separable convolutional layers were used as backbone networks for classification. Xing et al converted the 3D full image into 2D dynamic image, and then took the classical neural network and attention mechanism as the network model [30]. For these baseline methods, we maintain the parameter settings of the original paper. The results of evaluation indicators are shown in Table III. It can be found that our proposed method is the best in most indicators.

2) *Comparison with state-of-the-art methods:* In this section, we focus on comparing the classification performance of other widely used methods. The work investigated [31]–[37] and other MRI monomodal classification performance, see Table IV below. Multi-modality [32]–[34], [36], [38]–[41] includes MRI + PET, MRI + PET + biomarkers, MRI + DTI, and MRI + Cognitive scores, as shown in Table V.

TABLE III. COMPARISON OF PERFORMANCE OF DIFFERENT CLASSIFICATION METHODS

Methods	Type	AD vs. NC (%)			MCI vs. NC (%)		
		ACC	AUC	F1	ACC	AUC	F1
Korolev et al. [17]	3D	70.00	69.54	67.57	73.75	67.40	82.64
Spasov et al. [18]	3D+Cli Info	98.00	99.38	97.70	80.62	83.91	84.52
Xing et al. [30]	3D	86.25	85.93	86.08	78.13	70.25	84.58
Proposed	3D+2D+Cli Info	99.00	99.12	98.67	81.63	82.41	87.06

TABLE IV. PERFORMANCE COMPARISON OF DIFFERENT METHODS IN SINGLE-MODALITY STUDY

Study	Modality	Group			Method	Accuracy(%)	
		AD	MCI	NC		AD vs. NC	MCI vs. NC
Liu et al. [31]	MRI	97	234	128	multi-view: multi-template:SVM	93.83	-
Zhu et al. [32]	MRI	51	99	52	ROI-based:Relational regularization feature selection	93.70	79.70
Liu et al. [33]	MRI	93	204	100	Patch-based: Cascaded CNNs	84.97	-
Aderghal et al. [34]	MRI	188	399	228	Hippocampus:CNN; Transfer learning	90.00	72.50
Basaia et al. [35]	MRI	295	$510^a + 253^b$	352	Patch-based:CNN	99.20	76.10^c 87.10^d
Shao et al. [36]	MRI	160	460	160	Hypergraph; Multikernel SVM	88.30	69.14
Al-Khuzai et al. [37]	MRI	170	-	70	Slices-based:CNN	99.30	-

Group: $a^a + b^b$: the number of sMCI is a and the number of pMCI is b ; Accuracy: c^c : the accuracy of sMCI/NC is c ; d^d : the accuracy of pMCI/NC is d .

TABLE V. PERFORMANCE COMPARISON OF DIFFERENT METHODS IN MULTI-MODALITY STUDY

Study	Modality	Group			Method	Accuracy(%)	
		AD	MCI	NC		AD vs. NC	MCI vs. NC
Zhang et al. [38]	MRI+PET+CSF	51	99	52	ROI-based: Kernel combination	93.20	76.40
Zhu et al. [32]	MRI+PET	51	99	52	ROI-based:Relational regularization feature selection	95.70	79.90
Shi et al. [39]	MRI+PET	51	99	52	ROI-based: Multimodal SDPN	97.13	87.24
Liu et al. [33]	MRI+PET	93	204	100	Patch-based: Cascaded CNNs	93.26	74.34
Aderghal et al. [34]	MRI+DTI	48	108	58	Hippocampus; CNN; Transfer learning	92.50	80.00
Shi et al. [40]	MRI+PET+CSF	51	99	52	ROI-based: Coupled boosting; Coupled metric ensemble	94.85	79.88
Shao et al. [36]	MRI+PET	160	460	160	Hypergraph; Multi-kernel SVM	92.51	82.53
Hett et al. [41]	MRI+Cognitive scores	130	216	213	Patch-based: Multiscale graphs	91.60	-
Proposed	MRI+Cognitive scores	45	123	56	CNN:Transfer learning; Ensemble	99.00	81.63

CSF = Cerebrospinal fluid

For AD/NC classification, the accuracy of the single-modal methods listed in Table IV has classification results below 90.00%, while the accuracy of most multi-modal methods is above 90.00%. For MCI/NC classification, the accuracy of most single-modal methods is below 80.00%, while the accuracy of most multi-modal methods is above 80%. Among the listed studies, Zhu et al. [32], Liu et al. [33], Aderghal et al. [34] and Shao et al. [36] performed single-modal and multi-modal tests on the proposed method. The results show that, compared with single-modal data, the use of multi-modal data can obtain higher classification accuracy. In addition, our method achieves the best performance of 99.00% in the AD/NC classification and 81.63% in the MCI/NC classification with resnet18 as the backbone network. It is worth noting that, due to potential differences in data selection, preprocessing and even data set division, the results obtained by different methods are actually incomparable. The purpose of the comparison is only to provide an overview of other results and to show the baseline of existing methods.

3) *Different backbone neural networks:* We performed other backbone neural network training. To evaluate the classification performance of different backbone neural networks, we used a five-fold cross-validation strategy to calculate the classification performance. Specifically, the entire subject sample set was divided into five subsets equally, and the subject samples within one subset were selected as test samples each time, and all remaining subject samples within the other four subsets were used to train the classifiers. This process was repeated five times independently to avoid any bias introduced by randomly dividing the dataset in cross-validation. We take the average of the three experiments as the final result of the data. The results are shown in the Table VI and Table VII.

TABLE VI. PERFORMANCE OF DIFFERENT BACKBONE NEURAL NETWORKS FOR AD VS. NC (%)

Methods	ACC	AUC	F1	PRECISION	RECALL	AP
AMIM+resnet18	98.33	98.75	98.01	96.85	99.48	96.67
AMIM+resnet34	98.33	98.75	98.03	97.50	98.79	96.95
AMIM+resnet50	98.67	99.27	98.35	97.50	99.39	97.23
AMIM+resnet101	99.00	98.68	98.67	97.50	1.0	97.50
AMIM+vgg11	98.33	99.12	98.08	97.50	98.88	97.05
AMIM+googlenet	97.00	99.19	96.67	97.38	96.20	95.72
AMIM+alexnet	98.00	98.75	97.38	96.09	99.05	95.60

TABLE VII. PERFORMANCE OF DIFFERENT BACKBONE NEURAL NETWORKS FOR MCI VS. NC (%)

Methods	ACC	AUC	F1	PRECISION	RECALL	AP
AMIM+resnet18	80.51	80.73	86.72	84.52	95.42	81.86
AMIM+resnet34	82.52	82.78	88.18	86.51	96.85	83.65
AMIM+resnet50	78.61	79.98	85.30	87.18	95.76	83.10
AMIM+resnet101	78.40	80.33	85.54	86.52	96.15	81.66
AMIM+vgg11	80.09	80.74	86.97	82.44	97.95	80.83
AMIM+googlenet	79.71	77.80	86.24	84.02	97.11	81.45
AMIM+alexnet	82.86	82.71	88.19	85.57	93.90	83.84

V. DISCUSSION

A. Performance analysis of different views

3D MRI exists three views (axial, coronal and sagittal view). We used the axial view. The different views of the 3D global information and the maximum information slice are shown in Fig. 4. Pan et al. [42] proposed a Multi-view Separable Pyramid Network (MiSePyNet), in which representations are learned from axial, coronal and sagittal views of PET scans so as to offer complementary information and then combined to make a decision jointly. The experimental results show that the performance of the axial view is the best and multi-view fusion effect is better than the single-view. Next, this paper will discuss and analysis the classification performance of different views through experiments.

We performed experiments for different views by using the same parameter settings. The overall experimental results of AD versus NC and MCI versus NC are shown in Table VIII and Table IX. The classification performance of different views for each model has been given in the table. The best evaluation metrics are shown in bold. The experimental results show that AD versus NC and MCI versus NC experimental

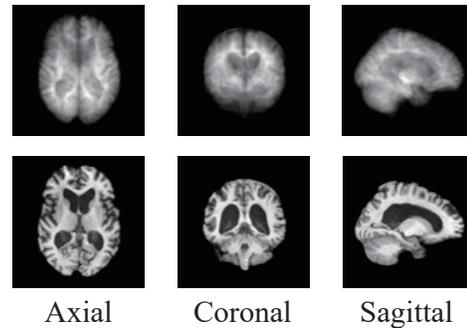


Fig. 4. Different views of the global information image and the maximum information slice

TABLE VIII. PERFORMANCE OF DIFFERENT VIEWS FOR AD VS. NC (%)

Type	Views	ACC	AUC	F1
Slice	Axial view	83.14	82.59	83.09
	Coronal view	85.24	87.49	82.39
	Sagittal view	81.19	86.23	79.41
Dynamic	Axial view	90.00	91.90	88.63
	Coronal view	86.05	87.49	86.44
	Sagittal view	89.05	93.82	88.58
Fusion	Axial view	92.00	93.62	91.06
	Coronal view	86.10	84.33	85.45
	Sagittal view	86.14	85.48	85.39

TABLE IX. PERFORMANCE OF DIFFERENT VIEWS FOR MCI VS. NC (%)

Type	Views	ACC	AUC	F1
Slice	Axial view	71.00	62.34	80.76
	Coronal view	72.10	67.71	81.35
	Sagittal view	69.89	63.97	80.89
Dynamic	Axial view	77.10	74.45	84.61
	Coronal view	72.65	66.63	82.23
	Sagittal view	72.10	62.72	82.66
Fusion	Axial view	74.29	72.39	82.26
	Coronal view	71.54	55.46	81.14
	Sagittal view	70.44	54.81	79.65

classification performance are similar. In the model with only slices as input, the overall evaluation metrics of coronal view are the best, the evaluation metrics of the other two views have small differences with them. In the model with only dynamic images as input, the AUC and F1 metrics of the axial view are the highest. In the whole hybrid model of dynamic image and slices, all evaluation metrics of the axial plane is the highest.

For the overall results obtained from these three different views, we can find that the classification performance of each view is not very different and the axial view is relatively better. In this paper, we only conducted experiments for one view. In the future, we will analyze three views of the image together. Different views show different information and how to obtain more comprehensive information plays a more important role

for Alzheimer's disease diagnostic.

B. Data Pairing

The dataset we used is from the "Spatially normalized, masked and N3-corrected T1 images" category in the ADNI public dataset. The dataset in this category contain paired MRI and clinical information. However, for the other categories of dataset in the ADNI, there were cases where subjects had not undergone the MMSE, ADAS11 examination. For patients with clinical information, we effectively combined the clinical information, which helped to improve the classification performance of the model. In the future, we will try to improve the diagnostic performance of ADNI with only some of the subjects' basic information (gender, age, etc.).

VI. CONCLUSION

In this paper, we proposed a multimodal adaptive weighted model, which takes global information images, maximum information slices and clinical information as multimodal inputs for the first time. Our model can effectively solve the problem of missing global information in slice classification. At the same time, the use of image information entropy selection slices can solve the subjective uncertainty of human selection. Using an adaptive weighting method to optimize the weights, it can combine the weights of different models more accurately than the grid search method. Our model achieves the best results in terms of classification performance, compared with the latest methods. The combination of medical images and clinical information for Alzheimer's disease classification is the future trend. Next, we will try to investigate how to better combine clinical information with medical images.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their help. This work was supported by the National Natural Science Foundation of China (Grant No. 62076044) ; the Key Project of Chongqing Natural Science Foundation in China (Grant No. cstc2019jcyjzdxmX0011); Chongqing Graduate Research Innovation Project of China (Grant No. CYS21307).

REFERENCES

- [1] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, and H. M. Arrighi, "Forecasting the global burden of Alzheimer's disease," *Alzheimer's & Dementia*, vol. 3, no. 3, pp. 186–191, 2007.
- [2] W. Jagust, "Vulnerable neural systems and the borderland of brain aging and neurodegeneration," *Neuron*, vol. 77, no. 2, pp. 219–234, 2013.
- [3] J. Cummings, G. Lee, A. Ritter, M. Sabbagh, and K. Zhong, "Alzheimer's disease drug development pipeline: 2019," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, vol. 5, no. 1, pp. 272–293, 2019.
- [4] A. Atri, "Current and future treatments in Alzheimer's disease," in *Seminars in neurology*, vol. 39, no. 02, 2019, pp. 227–240.
- [5] M. A. Ebrahimighanavieh, S. Luo, and R. Chiong, "Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review," *Computer Methods and Programs in Biomedicine*, vol. 187, p. 105242, 2020.
- [6] F. Li, L. Tran, K.-H. Thung, S. Ji, D. Shen, and J. Li, "A Robust Deep Model for Improved Classification of AD/MCI Patients," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 5, pp. 1610–1616, 2015.
- [7] H.-I. Suk, C.-Y. Wee, S.-W. Lee, and D. Shen, "State-space model with deep learning for functional dynamics estimation in resting-state fmri," *NeuroImage*, vol. 129, pp. 292–307, 2016.
- [8] Y. Li, F. Meng, and J. Shi, "Learning using privileged information improves neuroimaging-based cad of alzheimer's disease: a comparative study," *Medical & Biological Engineering & Computing*, vol. 57, pp. 1605–1616, 2019.
- [9] M. I. Razzak, S. Naz, and A. Zaib, *Deep Learning for Medical Image Processing: Overview, Challenges and the Future*, N. Dey, A. S. Ashour, and S. Borra, Eds. Cham: Springer International Publishing, 2018.
- [10] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [11] M. Hon and N. Khan, "Towards Alzheimer's disease classification through transfer learning," in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Los Alamitos, CA, USA: IEEE Computer Society, 2017, pp. 1166–1169.
- [12] J. Islam and Y. Zhang, "Early Diagnosis of Alzheimer's Disease: A Neuroimaging Study with Deep Learning Architectures," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018, pp. 1962–19622.
- [13] F. Zhang, B. Pan, P. Shao, P. Liu, S. Shen, P. Yao, and R. X. Xu, "A Single Model Deep Learning Approach for Alzheimer's Disease Diagnosis," *Neuroscience*, vol. 491, pp. 200–214, 2022.
- [14] S. Luo, X. Li, and J. Li, "Automatic Alzheimer's Disease Recognition from MRI Data Using Deep Learning Method," *Journal of Applied Mathematics and Physics*, vol. 05, pp. 1892–1898, 2017.
- [15] R. Jain, N. Jain, A. Aggarwal, and D. J. Hemanth, "Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images," *Cognitive Systems Research*, vol. 57, pp. 147–159, 2019.
- [16] J. Liu, M. Li, Y. Luo, S. Yang, W. Li, and Y. Bi, "Alzheimer's disease detection using depthwise separable convolutional neural networks," *Computer Methods and Programs in Biomedicine*, vol. 203, p. 106032, 2021.
- [17] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3D brain MRI classification," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017, pp. 835–838.
- [18] S. Spasov, L. Passamonti, A. Duggento, P. Liò, and N. Toschi, "A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease," *NeuroImage*, vol. 189, pp. 276–287, 2019.
- [19] D. Cheng and M. Liu, "Cnns based multi-modality classification for AD diagnosis," in *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2017, pp. 1–5.
- [20] M. Liu, D. Cheng, K. Wang, and Y. Wang, "Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis," *Neuroinformatics*, vol. 16, no. 3, pp. 295–308, 2018.
- [21] S. Qiu, P. S. Joshi, M. I. Miller, C. Xue, X. Zhou, C. Karjadi, G. H. Chang, A. S. Joshi, B. Dwyer, S. Zhu, M. Kaku, Y. Zhou, Y. J. Alderazi, A. Swaminathan, S. Kedar, M.-H. Saint-Hilaire, S. H. Auerbach, J. Yuan, E. A. Sartor, R. Au, and V. B. Kolachalama, "Development and validation of an interpretable deep learning framework for Alzheimer's disease classification," *Brain*, vol. 143, no. 6, pp. 1920–1933, 2020.
- [22] J. M. Rondina, L. K. Ferreira, F. L. de Souza Duran, R. Kubo, C. R. Ono, C. C. Leite, J. Smid, R. Nitrini, C. A. Buchpiguel, and G. F. Busatto, "Selecting the most relevant brain regions to discriminate Alzheimer's disease patients from healthy controls using multiple kernel learning: A comparison across functional and structural imaging modalities and atlases," *NeuroImage: Clinical*, vol. 17, pp. 628–641, 2018.
- [23] B. Duraisamy, J. V. Shanmugam, and J. Annamalai, "Alzheimer disease detection from structural MR images using FCM based weighted probabilistic neural network," *Brain imaging and behavior*, vol. 13, no. 1, pp. 87–110, 2019.
- [24] S. Kornblith, J. Shlens, and Q. V. Le, "Do better imagenet models transfer better?" in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 2656–2666.

- [25] A. Bobick and J. Davis, "The recognition of human movement using temporal templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 257–267, 2001.
- [26] B. Fernando, E. Gavves, M. José Oramas, A. Ghodrati, and T. Tuytelaars, "Modeling video evolution for action recognition," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 5378–5387.
- [27] H. Bilen, B. Fernando, E. Gavves, A. Vedaldi, and S. Gould, "Dynamic Image Networks for Action Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 3034–3042.
- [28] S. Qiu, G. Chang, M. Panagia, D. Gopal, R. Au, and V. Kolachalama, "Fusion of deep learning models of MRI scans, Mini-Mental State Examination, and logical memory test enhances diagnosis of mild cognitive impairment," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 10, 2018.
- [29] C. Studholme, D. Hill, and D. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognition*, vol. 32, no. 1, pp. 71–86, 1999.
- [30] X. Xing, G. Liang, H. Blanton, M. U. Rafique, C. Wang, A.-L. Lin, and N. Jacobs, "Dynamic Image for 3D MRI Image Alzheimer's Disease Classification," in *Computer Vision – ECCV 2020 Workshops*, A. Bartoli and A. Fusiello, Eds. Cham: Springer International Publishing, 2020, pp. 355–364.
- [31] M. Liu, D. Zhang, E. Adeli, and D. Shen, "Inherent Structure-Based Multiview Learning With Multitemplate Feature Representation for Alzheimer's Disease Diagnosis," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1473–1482, 2016.
- [32] X. Zhu, H.-I. Suk, L. Wang, S.-W. Lee, and D. Shen, "A novel relational regularization feature selection method for joint regression and classification in AD diagnosis," *Medical Image Analysis*, vol. 38, pp. 205–214, 2017.
- [33] M. Liu, D. Cheng, K. Wang, and Y. Wang, "Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis," *Neuroinformatics*, vol. 16, no. 3, pp. 295–308, 2018.
- [34] K. Aderghal, A. Khvostikov, A. Krylov, J. Benois-Pineau, K. Afdel, and G. Catheline, "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.
- [35] S. Basaia, F. Agosta, L. Wagner, E. Canu, G. Magnani, R. Santangelo, and M. Filippi, "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks," *NeuroImage: Clinical*, vol. 21, p. 101645, 2019.
- [36] W. Shao, Y. Peng, C. Zu, M. Wang, and D. Zhang, "Hypergraph based multi-task feature selection for multimodal classification of Alzheimer's disease," *Computerized Medical Imaging and Graphics*, vol. 80, p. 101663, 2020.
- [37] F. E. Al-Khuzai, O. Bayat, and A. D. Duru, "Diagnosis of Alzheimer disease using 2D MRI slices by convolutional neural network," *Applied Bionics and Biomechanics*, vol. 2021, 2021.
- [38] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *NeuroImage*, vol. 55, no. 3, pp. 856–867, 2011.
- [39] J. Shi, X. Zheng, Y. Li, Q. Zhang, and S. Ying, "Multimodal Neuroimaging Feature Learning With Multimodal Stacked Deep Polynomial Networks for Diagnosis of Alzheimer's Disease," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, pp. 173–183, 2018.
- [40] Y. Shi, H.-I. Suk, Y. Gao, S.-W. Lee, and D. Shen, "Leveraging Coupled Interaction for Multimodal Alzheimer's Disease Diagnosis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 1, pp. 186–200, 2019.
- [41] K. Hett, V.-T. Ta, I. Oguz, J. V. Manjón, and P. Coupé, "Multi-scale graph-based grading for Alzheimer's disease prediction," *Medical Image Analysis*, vol. 67, p. 101850, 2021.
- [42] e. a. Pan X, Phan T L, "Multi-view separable pyramid network for ad prediction at mci stage by 18 f-fdg brain pet imaging," *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 81–92, 2020.