

Article

Diagnosis of Alzheimer's Disease with Ensemble Learning Classifier and 3D Convolutional Neural Network

Peng Zhang, Shukuan Lin *, Jianzhong Qiao and Yue Tu

Department of Computer Science and Engineering, Northeastern University, Shenyang 110819, China; zhangpeng.coder@gmail.com (P.Z.); qiaojianzhong@cse.neu.edu.cn (J.Q.); 1810615@stu.neu.edu.cn (Y.T.)

* Correspondence: linshukuan@cse.neu.edu.cn

Abstract: Alzheimer's disease (AD), the most common type of dementia, is a progressive disease beginning with mild memory loss, possibly leading to loss of the ability to carry on a conversation and respond to environments. It can seriously affect a person's ability to carry out daily activities. Therefore, early diagnosis of AD is conducive to better treatment and avoiding further deterioration of the disease. Magnetic resonance imaging (MRI) has become the main tool for humans to study brain tissues. It can clearly reflect the internal structure of a brain and plays an important role in the diagnosis of Alzheimer's disease. MRI data is widely used for disease diagnosis. In this paper, based on MRI data, a method combining a 3D convolutional neural network and ensemble learning is proposed to improve the diagnosis accuracy. Then, a data denoising module is proposed to reduce boundary noise. The experimental results on ADNI dataset demonstrate that the model proposed in this paper improves the training speed of the neural network and achieves 95.2% accuracy in AD vs. NC (normal control) task and 77.8% accuracy in sMCI (stable mild cognitive impairment) vs. pMCI (progressive mild cognitive impairment) task in the diagnosis of Alzheimer's disease.



Citation: Zhang, P.; Lin, S.; Qiao, J.; Tu, Y. Diagnosis of Alzheimer's Disease with Ensemble Learning Classifier and 3D Convolutional Neural Network. *Sensors* **2021**, *21*, 7634. <https://doi.org/10.3390/s21227634>

Academic Editor: Steve Ling

Received: 8 October 2021

Accepted: 15 November 2021

Published: 17 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: Alzheimer's disease; convolutional neural network; ensemble learning; data denoising

1. Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disease with insidious onset. Clinically, it is characterized by general dementia such as memory impairment, aphasia, impairment of visuospatial skills, and executive dysfunction. A global Alzheimer's disease report from the International Association for Alzheimer's disease pointed out that the number of AD patients in the world has increased dramatically. It may increase from 47 million in 2015 to 130 million in 2050, and the cost of AD treatment will also increase sharply, from 800 billion in 2015 to 2 trillion in 2030. This may bring a serious economic burden to the world. The elderly have a high incidence of Alzheimer's disease, and the aging of the population has become a serious social problem. Therefore, the elderly population with Alzheimer's disease will increase. Computer-aided diagnosis (CAD) of AD can help to detect the disease as early as possible and inhibit its progression, which has important significance in research and practical applications.

Mild cognitive impairment (MCI) is a prodromal state of dementia, which is the most important and suitable stage for AD preventive treatment. MCI is divided into progressive mild cognitive impairment (pMCI) and stable mild cognitive impairment (sMCI), where pMCI will eventually deteriorate into AD within 36 months, and sMCI is in a relatively stable condition with no lesions.

Magnetic resonance imaging (MRI) is a main medical neuroimaging method, which is an important basis for the diagnosis of Alzheimer's disease. Compared with computed tomography (CT), the organization structure of MRI is clearer. Brain morphometric pattern analysis using MRI data are proven to be effective in identifying anatomical differences between populations of AD patients and normal controls (NC). In recent years, deep learning has achieved significant results in the field of computer vision, especially in the

use of convolutional neural networks (CNNs) [1] to extract image features, which provides new ideas for AD diagnosis [2]. Many researchers have built 3D CNN models for AD diagnosis based on 3D MRI data, which contains more information compared with 2D data. However, 3D CNNs need a lot of training time. Some existing methods, for example, the voxel-based ones [3,4], patch-based ones [5,6] and ROI-based ones [7,8], tried to use part of an image for feature extraction [9–11], which require good priori knowledge to know which part affects the diagnosis of Alzheimer's disease. Because the etiology and pathology of Alzheimer's disease are not clear, these methods may reduce diagnosis accuracy. Some methods used whole images for feature extraction [12], which greatly increases calculation. Besides, most of the existing methods employed a single classifier for AD diagnosis, which limits the further improvement of diagnosis performance. Moreover, a limited number of MRI images were not fully used for AD diagnosis in current models. Especially, there are few studies on hard sample processing to improve the diagnosis accuracy of Alzheimer's disease. Therefore, although the existing methods have achieved good diagnostic results, there are still some problems that remain to be solved. The diagnosis of Alzheimer's disease has the following challenges:

Challenge 1: Although 3D MRI data contain more information than 2D data, the 3D MRI image-based 3D CNN models need more training time. Therefore, how to reduce the calculation in 3D CNN modeling without losing much useful information is a challenge for AD diagnosis. By observing 3D MRI images, we found that there are many noise regions near their edges. These noise regions contain a large number of useless boundary voxels. The introduction of a large amount of noise not only slows down the training speed, but also reduces diagnostic accuracy. Therefore, how to process the noise regions to increase training speed while improving diagnostic accuracy is a problem to be solved when we build our 3D CNN diagnosis model.

Challenge 2: For 3D MRI images, a single classifier limits the performance improvement of AD diagnosis. Meanwhile, inefficient utilization of a limited number of MRI images will lead to overfitting in the 3D CNN diagnosis models while hard samples usually reduce the accuracies of the models. Therefore, how to break through the limitation of using a single classifier and process a hard sample while fully utilizing the limited MRI images to improve diagnostic performance is another challenge. In our study, for this challenge, how to construct the corresponding 3D CNN network architecture and learn hard samples is a problem to be solved.

To tackle the above problems, we propose a novel diagnostic model with a data denoising module and diagnosis network module. The major contributions of this paper can be summarized as follows.

1. To process the noise regions to increase training speed while improving diagnostic accuracy, we propose a data denoising method that reduces the noises at the boundaries of the 3D MRI images through the designed clipping algorithm. By analyzing the differences between the images of different label groups, the difference image is obtained, and the image boundary is cropped according to the normalized mask threshold to reduce noises. The data denoising method can remove about 47% of useless information from the data for AD diagnosis, which not only speeds up the training process but also improves the accuracy of diagnosis.
2. Considering the limitation of a single classifier, the problem of hard sample learning and the issue of insufficient sample utilization, we propose a diagnosis network with multiple classifiers and serialization learning, where samples are repeatedly used in various classifiers with different weights according to their prediction errors in the previous classifier. Hard samples are assigned larger weights due to their larger prediction errors so that we can focus on them. The diagnosis network is trained based on the ensemble learning method with an improved loss function and adaptive fusion.
3. The experimental results demonstrate that our proposed method outperforms the state-of-the-art methods regarding both accuracy and efficiency. Specifically, our

proposed method has achieved a high accuracy 95.2% (AD vs. NC) and 77.8% (sMCI vs. pMCI).

The rest of this paper is organized as follows. We briefly introduce the most relevant studies in Section 2. In Section 3, we describe the data used in this study and present our method in detail. In Section 4, we show experimental results and performance evaluation. Finally, the conclusions are given in Section 5.

2. Related Work

A key component of an MRI-based AD diagnosis system is to determine how to extract features from MRI images. Generally speaking, the feature extraction of existing AD/MCI diagnosis can be roughly divided into four categories, including voxel-based feature extraction, patch-based feature extraction, region-of-interest (ROI) based feature extraction, and whole-image-based feature extraction.

More specifically, the voxel-based features measure the local tissues associated with AD [13]. Baron et al. [3] proposed a feature extraction method based on voxels and utilized the statistical features of voxels for diagnosis. Li et al. [4] used the voxel information of the cerebral cortex to extract features and input them into a support vector machine (SVM) [14] for diagnosis. Compared with a relatively small number of images (such as tens or hundreds) used for model training, the feature dimension of voxels is very high (such as millions). This kind of method usually faces the challenge of overfitting.

The patch-based methods use all or part of patches as inputs for diagnosis [15]. Zhang et al. [5] used the regression forest algorithm to predict the locations of AD lesions, then cropped MRI images according to these locations, and finally input them to SVM for diagnosis. Liu et al. [6] used positron emission computed tomography (PET) and MRI images for multi-modal diagnosis. Besides, they divided each MRI and PET image into patches according to corresponding positions and input to a cascade network. Liu et al. [16] identified discriminative anatomical landmarks from each MRI image via a data-driven strategy and extracted multiple patches around these landmarks, which are input into multiple classification networks for multi-instance diagnosis learning. Pan et al. [17] used generative adversarial networks (GAN) to generate missing PET data, and then divided an image into multiple patches to input to a multi-instance neural network for learning. Suk et al. [18] used multi-modal data and divided them into multiple patches to input into the deep network for fusion diagnosis. Lian et al. [19] pre-determined the location of each lesion to generate the corresponding patch, and input them to CNN. Ortiz et al. [20] split gray matter images from each brain area into 3D patches according to the regions defined by the automated anatomical labeling atlas and these patches are used to train different deep belief networks. Although these patch-based methods can reduce the number of network parameters, the information between patches is relatively independent, which affects the diagnostic effectiveness.

The ROI-based methods mainly extract features from pre-segmented brain regions and construct classifiers [21]. Zeng et al. [7] used the gray matter regions in a brain to extract features and proposed an overall framework for optimizing SVM parameters. Wang et al. [8] used brain extraction tools to select the slices containing the hippocampus and input them into a convolutional neural network for diagnosis. Cui et al. [9] used gray matter in a brain to extract features and proposed a classification framework combining convolutional neural networks and recurrent neural networks for longitudinal analysis of structural MRI in AD diagnosis. Ju et al. [10] used functional MRI to divide a brain into 90 different regions for AD diagnosis. Cui et al. [11] diagnosed the disease by analyzing the characteristics of the shape and volume of biomarkers such as the hippocampus. These ROI-based methods can achieve good diagnostic results if the regions are selected correctly. However, region selection is based on prior knowledge, so all possible pathological locations in an entire brain may not be covered.

In the whole-image-based methods, an image is input to a network as a whole [22]. Liu et al. [23] cut 3D MRI data into 2D slices in three directions and input them into CNN

and recurrent neural network for fusion diagnosis. Wang et al. [12] input 3D MRI into DenseNet [24] for feature extraction for the first time, and then applied integrated learning for diagnosis. In this paper, five diagnostic subnetworks are designed and a new integrated diagnosis method is proposed. Islam et al. [25] used deep convolutional neural networks to train on a small dataset, and proposed an effective method for training deep learning models with unbalanced data. Maqsood et al. [26] utilized transfer learning to classify images by fine-tuning a pre-trained convolutional network, AlexNet. Karasawa et al. [27] designed a 3D CNN architecture for diagnosing AD and achieved good performance. Basaia et al. [28] augmented training data by deforming, cropping, rotating, flipping and scaling MRI data at different angles, and then input them into a convolutional neural network for diagnosis. Although this kind of method uses all the information of an image, there are some boundary noises, which make calculation expensive. There are also some methods that use other personal information besides images, such as age, gender, and checklist scores. For instance, Duc et al. [29] used a 3D convolutional network to jointly predict patients' diagnosis categories and their mental status scores. Feng et al. [30] designed a new deep learning framework. Specifically, the framework makes use of the advantages of 3D CNN and the full stacked bidirectional long short term memory network. Fsb LSTM is used to extract the hidden spatial information in deep feature map to further improve its performance. Although this method makes use of all the information of the image, because the boundary of MRI image has useless information, this calculation cost is greater and unnecessary for the network using 3D MRI.

The existing methods have the following problems: (1) there are a large number of voxels in a 3D MRI image with a large amount of useless boundary noise information, which not only affects the efficiency of 3D CNN modelling but also the accuracy of AD diagnosis. In current works, the useless boundary noise information did not be processed effectively. (2) Most of the existing methods used a single classifier for AD diagnosis, which limits the improvement of diagnosis performance. (3) In existing methods, a limited number of 3D MRI images were not sufficiently utilized for network training, while inputting 3D MRI samples into a network directly did not consider their different importance for diagnosis results. (4) Few existing methods focused on hard sample learning, leading to suboptimal diagnosis results. To tackle the above issues, we first propose a novel denoising method for 3D MRI data, which removes useless noise information. Then, we propose a diagnosis network, which contains multiple 3D CNNs with the corresponding improved loss functions as classifiers and the limited 3D MRI samples are repeatedly utilized in various classifiers with different weights. Especially, by incorporating the idea of AdaBoost [31], the weights of samples can be learned according to their prediction errors in the previous classifier, and hard samples are assigned larger weights due to their larger prediction errors so that we can improve diagnosis performance by focusing on hard sample learning.

3. Our Diagnosis Model

3.1. Model Architecture

As shown in Figure 1, our diagnosis model consists of two major parts, namely, the data denoising module and the diagnosis network module. In Figure 1, the green part represents the weights of the samples, and the color depth represents the weight size. After 3D MRI samples are cropped by the data denoising module, they are input into the diagnosis network containing multiple classifiers. In the training process, the weight of each sample is continually adjusted according to the diagnosis error of the corresponding classifier, and each sample is input to the next classifier together with its updated weight so that hard samples can be assigned high weights in the next learning. Meanwhile, the fusion weight of each classifier is adaptively learned so that the diagnosis results of all classifiers can be adaptively fused in the testing process.

More details about the data denoising module are presented in Section 3.2. In Sections 3.3 and 3.4, we introduce the diagnosis network including the adaptive fusion and the hard sample learning.

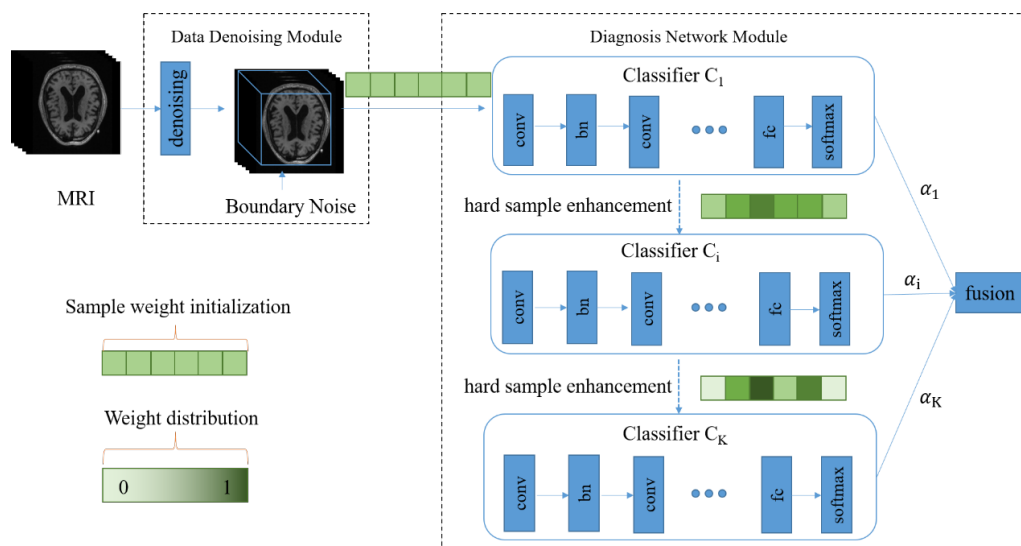


Figure 1. The architecture of the proposed diagnosis model for Alzheimer’s disease. There are two main components, including the data denoising module and diagnosis network module.

3.2. Data Denoising Module

The MRI images are mostly 3D data, which contain more information than 2D data while requiring a larger amount of calculation during training, which makes neural networks less efficient when training 3D MRI data. In Figure 2, we show the sectional views of a 3D MRI image along three directions (sagittal plane, coronal plane, and axial plane). We can clearly see that there are many black noise regions near the margins of each image. These regions contain a lot of useless boundary voxels, which have no impact on brain tissues and need to be processed prior to training the neural network. The introduction of a large amount of noise not only slows down the process of model training, but also reduces the accuracy of diagnosis. The motivation of the data denoising module is to provide more valuable data for AD diagnosis as the inputs of our neural network, so as to improve the diagnostic accuracy of our network model. Based on this, a data denoising method is proposed, which removes the marginal noise regions in raw data to accelerate network training and improve diagnostic accuracy.

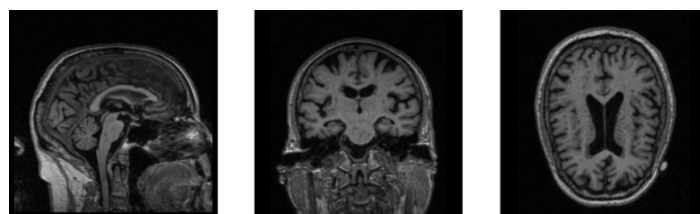


Figure 2. The sectional views of a 3D MRI image along three directions. The three views from left to right are sagittal, coronal, and axial.

In Figure 3, we show the data changes before and after data denoising. Through data denoising, the noise regions in the raw 3D image are cut out, and the size of the data input into the neural network can be reduced from $192 \times 192 \times 160$ to $160 \times 160 \times 120$. Because the noise data is reduced, the neural network model can better extract useful features for diagnosis. In addition, data denoising helps to reduce calculation by decreasing parameters of the fully connected layer in our diagnosis network.

In order to better understand the method proposed in this paper, the related concepts are defined as follows.

Definition 1. (*x*-Sample image): Given a dataset M of 3D MRI images where there are a total of four types of label values AD , NC , $sMCI$ and $pMCI$ for all sample images and the label value set

$L=\{AD, NC, sMCI, pMCI\}$, the sample image with the label $x \in L$ in M is called x -sample image, denoted by s^x .

Definition 2. (Image group): Given the dataset M , the set $G^x = \{s_1^x, s_2^x, \dots, s_m^x\} \subseteq M$, where $s_i^x (1 \leq i \leq m)$ is an x -sample image and m is the number of x -sample images in M , is called x -image group.

Definition 3. (Mean image): Given the x -image group $G^x = \{s_1^x, s_2^x, \dots, s_m^x\} \subseteq M$, the mean image $\overline{G^x}$ of G^x is defined as the image with the average voxel level of G^x . Each voxel value $\overline{G^x}(a, b, c)$ is calculated as follows:

$$\overline{G^x}(a, b, c) = \frac{1}{m} \sum_{i=1}^m s_i^x(a, b, c), \quad (1)$$

where $s_i^x(a, b, c)$ is the voxel value of the 3D sample image s_i^x .

Definition 4. (Difference image): Given the mean images $\overline{G^{x_1}}$ and $\overline{G^{x_2}}$, the difference image of $\overline{G^{x_1}}$ and $\overline{G^{x_2}}$ is defined as the image with their voxel difference at each position, denoted by $D^{x_1-x_2}$. Each voxel value $D^{x_1-x_2}(a, b, c)$ is calculated as follows.

$$D^{x_1-x_2}(a, b, c) = \overline{G^{x_1}}(a, b, c) - \overline{G^{x_2}}(a, b, c). \quad (2)$$

We use the AD-image group G^{AD} and the NC-image group G^{NC} to explain our data denoising process. First, we obtain their mean images $\overline{G^{AD}}$ and $\overline{G^{NC}}$ according to Equation (1). Then the difference image D^{AD-NC} is calculated according to Equation (2). By normalization, we can get the difference image \check{D}^{AD-NC} with the voxel values between 0 and 1. Next we transform \check{D}^{AD-NC} into its corresponding mask according to a given threshold α . Finally, each AD-sample image and each NC-sample image are cropped according to the mask to achieve noise reduction. In this way, the lesion areas corresponding to large differences in an image will be preserved, and the useless voxels can be cut off. In Figure 3, we show the views of a raw image at three dimensions and their corresponding denoised views. It can be found that the useless information near the boundaries is reduced, but the main information is not lost. The implementation details of data denoising are shown in Algorithm 1.

Algorithm 1 Data denoising process

Input: AD-Image group $G^{AD} = \{s_1^{AD}, s_2^{AD}, \dots, s_m^{AD}\}$, NC-Image group $G^{NC} = \{s_1^{NC}, s_2^{NC}, \dots, s_n^{NC}\}$, the threshold α **Output:** The set R of the positions where the voxel values will be preserved in denoising

- 1: Calculate the mean images $\overline{G^{AD}}$ and $\overline{G^{NC}}$
 - 2: Calculate the difference image D^{AD-NC} of $\overline{G^{AD}}$ and $\overline{G^{NC}}$
 - 3: Find the maximal value max in D^{AD-NC}
 - 4: Each value of D^{AD-NC} is divided by max to normalize it and obtain \check{D}^{AD-NC}
 - 5: **for** each position p in \check{D}^{AD-NC} **do**
 - 6: **if** $value_p > \alpha$ **then**
 - 7: $mask_p = 1$
 - 8: **else**
 - 9: $mask_p = 0$
 - 10: **end if**
 - 11: **end for**
 - 12: Add the coordinates with the value 1 to the set R
 - 13: **return** R
-

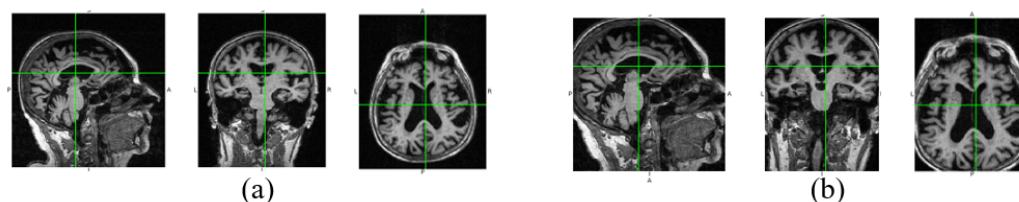


Figure 3. Raw image and the denoised image. The three views of the raw image are shown in (a), and (b) is the corresponding denoised views.

3.3. Diagnosis Network Module

As we know, medical images available for training are less adequate than ordinary images, so it is more necessary to make full use of the information of each sample in the training process. Especially for 3D medical images, each sample has a lot of voxel points and contains a lot of information. Therefore, the training process should focus on learning misclassified samples. On the one hand, we utilize multiple classifiers to repeatedly train limited MRI images so that they can be fully exploited. On the other hand, we continually adjust the weight of each sample in each classifier according to its prediction error in the previous classifier to punish misclassified samples and improve the accuracy of diagnosis.

3.3.1. The Structure of Diagnosis Network

The data used in this paper are 3D images. Therefore, we build a 3D convolutional diagnosis network so that the correlation between different dimensions is not lost, as shown in Figure 1. The diagnosis network incorporates the idea of ensemble learning and utilizes multiple CNNs as classifiers to achieve adaptive fusion. In order to reduce network parameters, inspired by VGG [32], we design $3 \times 3 \times 3$ small convolution kernels to extract image features in each classifier. We use ReLU [33] as the activation function. ReLU activation function is widely used in DNNs (Deep Neural Networks), especially in CNNs (Convolutional Neural Networks) because of its high efficiency, quick convergence, and gradient vanishing prevention. In our study, the input data are 3D MRI images with high dimensions. Accordingly, our diagnosis network is 3D CNN. It is necessary to use such an efficient and effective activation function. Then, we add a BN layer [34] after each convolutional layer to reduce data fluctuations and accelerate training speed. MaxPooling [35] is executed to increase receptive fields. In order to avoid overfitting, a dropout [36] is used in all fully connected layers. Finally a softmax layer is used for prediction. The specific network settings are shown in Table 1.

Table 1. Basic setting of each classifier.

Layer	Operation	Size	Input	Output
0	conv(bn)	$3 \times 3 \times 3 \times 8$	$192 \times 192 \times 160 \times 1$	$192 \times 192 \times 160 \times 8$
1	maxpooling	$2 \times 2 \times 2$	$192 \times 192 \times 160 \times 8$	$96 \times 96 \times 80 \times 8$
2	conv(bn)	$3 \times 3 \times 3 \times 16$	$96 \times 96 \times 80 \times 8$	$96 \times 96 \times 80 \times 16$
3	maxpooling	$2 \times 2 \times 2$	$96 \times 96 \times 80 \times 16$	$48 \times 48 \times 40 \times 16$
4	conv(bn)	$3 \times 3 \times 3 \times 32$	$48 \times 48 \times 40 \times 16$	$48 \times 48 \times 40 \times 32$
5	conv(bn)	$3 \times 3 \times 3 \times 32$	$48 \times 48 \times 40 \times 32$	$48 \times 48 \times 40 \times 32$
6	maxpooling	$2 \times 2 \times 2$	$48 \times 48 \times 40 \times 32$	$24 \times 24 \times 20 \times 32$
7	conv(bn)	$3 \times 3 \times 3 \times 64$	$24 \times 24 \times 20 \times 32$	$24 \times 24 \times 20 \times 64$
8	conv(bn)	$3 \times 3 \times 3 \times 64$	$24 \times 24 \times 20 \times 64$	$24 \times 24 \times 20 \times 64$
9	maxpooling	$2 \times 2 \times 2$	$24 \times 24 \times 20 \times 64$	$12 \times 12 \times 10 \times 64$
10	conv(bn)	$3 \times 3 \times 3 \times 64$	$12 \times 12 \times 10 \times 64$	$12 \times 12 \times 10 \times 64$
11	conv(bn)	$3 \times 3 \times 3 \times 64$	$12 \times 12 \times 10 \times 64$	$12 \times 12 \times 10 \times 64$
12	maxpooling	$2 \times 2 \times 2$	$12 \times 12 \times 10 \times 64$	$6 \times 6 \times 5 \times 64$
13	fc	2048	$6 \times 6 \times 5 \times 64$	2048
14	fc	2048	2048	2048
15	fc	2	2048	2

3.3.2. The Training of the Diagnosis Network

We use $D_k = (w_1^k, w_2^k, \dots, w_N^k)$ to represent the weight vector of samples in the k th classifier, where N is the number of samples and $w_i^k (1 \leq i \leq N)$ is the weight of the i th sample in the k th classifier. The sample weights in the first classifier are initialized as:

$$D_1 = (w_1^1, \dots, w_i^1, \dots, w_N^1), w_i^1 = \frac{1}{N}, i = 1, 2, \dots, N. \quad (3)$$

In order to adjust the weight of each sample in next classifier and achieve adaptive fusion, the error of each classifier is defined as follows:

$$e_k = \sum_{i=1}^N w_i^k I(C_k(x_i), y_i), \quad (4)$$

where e_k represents the error of the k th classifier, and $C_k(x_i)$ is the output corresponding to the sample x_i in the k th classifier. y_i represents the ground truth label of x_i , and $I(\cdot)$ is an indicator function, as shown in Equation (5).

$$I(C_k(x_i), y_i) = \begin{cases} 1 & C_k(x_i) \neq y_i \\ 0 & C_k(x_i) = y_i. \end{cases} \quad (5)$$

By the error in Equation (4), we can learn an adaptive fusion weight for each classifier as follows:

$$\alpha_k = \frac{1}{2} \log \frac{1 - e_k}{e_k} \quad (6)$$

where α_k represents the fusion weight of the k th classifier. According to Equation (6), the classifier with a larger error will be assigned a smaller fusion weight and vice versa. According to the error e_k of the k th classifier, the weight w_i^{k+1} of each sample x_i in the $(k + 1)$ th classifier is adjusted as:

$$w_i^{k+1} = \frac{w_i^k}{Z_k} \exp(-\alpha_k y_i C_k(x_i)), i = 1, 2, \dots, N \quad (7)$$

$$Z_k = \sum_{i=1}^N w_i^k \exp(-\alpha_k y_i C_k(x_i)), \quad (8)$$

where Z_k is the normalization factor of the k th classifier. In the k th classifier, if x_i is classified incorrectly, that is, $C_k(x_i)$ and y_i are inconsistent, its weight will increase in the $(k + 1)$ th classifier.

3.3.3. Loss Function

The cross-entropy loss is frequently used for image classification. However, the traditional cross-entropy loss does not consider the difference between samples. In this paper, we improve the traditional cross-entropy loss by incorporating the weight of each sample, as shown in Equation (9). From the training process of the diagnosis network, we can know that the weight of each sample may be adjusted in different classifiers. So each classifier has a separate loss function w.r.t sample weights.

$$\begin{aligned} loss^k = & \frac{1}{N} \sum_{i=1}^N -w_i^k [y_i \cdot \log(C_k(x_i)) \\ & + (1 - y_i) \cdot \log(1 - C_k(x_i))] \end{aligned} \quad (9)$$

where $loss^k$ is the loss function of the k th classifier.

3.4. Fusion Diagnosis

In the training process of each classifier, its error is calculated independently by Equation (4), and the fusion weight of each classifier in diagnosis is determined by its error as shown in Equation (6). The commonly used ensemble learning prediction method is based on voting. Each classifier votes on the prediction result, and the prediction with the most votes is the final result. However, if some classifiers are not trained accurately, their prediction results are unreliable. Therefore, it is necessary to assign a higher weight to a classifier with less error in the fusion diagnosis. In this paper, we adaptively fuse the results of various classifiers according to their errors as follows:

$$f(x) = \sum_{k=1}^K \alpha_k C_k(x), \quad (10)$$

where f is our fusion diagnosis function. x is a denoised testing sample, and K is the number of classifiers.

4. Experiments

4.1. Dataset

We conduct experiments on an ADNI dataset from Alzheimer's Disease Neuroimaging Initiative [37]. The data are 3D MRI images. The imaging protocol parameters are shown in Table 2. The image format is neuroimaging informatics technology initiative (NIFTI).

Table 2. Setting of the imaging protocol. The left is the imaging parameter, and the right is the corresponding value.

Protocol Parameter	Value
Acquisition Plane	Sagittal
Acquisition Type	3D
Field Strength	1.5 tesla
Slice Thickness	1.2 mm
TE	3.5–3.7 ms
TI	1000.0 ms
TR	3000.0 ms
Weighting	T1

The ADNI data distribution is shown in Figure 4. There are a total of 1461 MRI images, including 767 males and 694 females, aged between 54 and 96 years. We conduct training and testing on the ADNI dataset to verify the effectiveness of the data denoising module and diagnosis network module proposed in this paper. The ratio of training set to the testing set is 7:3. In order to be consistent with the experimental results of existing work for comparison, we conduct AD vs. NC and sMCI vs. pMCI diagnostic experiments, in which the sMCI-image group and pMCI-image group adopt the same denoising method as AD-image group and NC-image group.

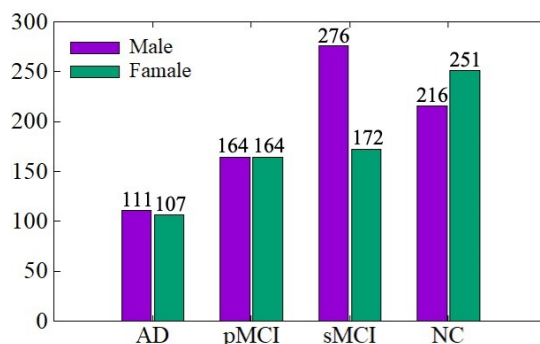


Figure 4. ADNI data distribution.

4.2. Experiment Setting

4.2.1. Implement Details

The experiments are implemented on a computer with a single GPU (NVIDIA RTX 2080 Ti) and the platform of Pytorch. To accelerate the training process and avoid local minimums, we use the Adam algorithm [38] as the training optimizer. The hyperparameters of our network are as follows: the batchsize is set to 4, and the epoch is set to 50. The initial learning rate equals to 10^{-4} , and the decay rate is 10^{-4} .

4.2.2. Performance Evaluation

In our diagnosis tasks, the predicted results are divided into the following four situations: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Among them, TP indicates that a positive sample is predicted to be a positive sample. TN indicates that a negative sample is predicted to be a negative sample. FP indicates that a negative sample is predicted to be a positive sample. FN indicates that a positive sample is predicted to be a negative sample. We employ the following four indicators to evaluate our diagnosis model: Accuracy, Specificity, Sensitivity, and Area under curve (AUC). Accuracy is the proportion of correctly diagnosed samples among the whole testing samples, as shown in Equation (11).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

Specificity represents the percentage of correctly diagnosed samples in disease-free testing samples, as shown in Equation (12).

$$Specificity = \frac{TN}{TN + FP} \quad (12)$$

Sensitivity reflects the ability of a model to recognize Alzheimer's patients in all positive samples, as shown in Equation (13).

$$Sensitivity = \frac{TP}{TP + FN} \quad (13)$$

AUC is obtained by summing the area under a receiver operating characteristic (ROC) curve. The ROC curve expresses the sensitivity of a classifier to the distance between positive and negative samples, but the curve cannot clearly show which classifier is better. So, AUC is used as the numerical indicator for comparison. The larger the AUC, the more accurate the corresponding model.

4.3. Experimental Results

4.3.1. Parameter Analysis

Analysis of the threshold α in the data denoising module. As mentioned in Section 3.2, we first perform data denoising before inputting samples into the diagnosis network. The α value will influence the denoising effect, thereby affecting the diagnostic accuracy. The curve of accuracy versus α is shown in Figure 5 and the accuracy reaches a peak at $\alpha = 0.5$ in each diagnostic task (i.e., AD vs. NC and sMCI vs. pMCI). When α is less than 0.5, the boundary noise of images is less cropped and the effect is not ideal. On the contrary, when α is larger than 0.5, some useful information may be cropped. Thereby, we set $\alpha = 0.5$ in our experiments.

Analysis of the number of classifiers. We incorporate the idea of ensemble learning in our diagnosis network and train multiple classifiers for fusion diagnosis. The diagnosis effect is improved by increasing the loss of hard samples and reducing the weights of the classifiers with high errors. We show the accuracy of our model with the number of classifiers K in Figure 6. When K is small, the diagnostic effect is improved with K increasing. When K is greater than 4, the performance is basically unchanged, so the number of classifiers K is set to 4 in our experiments.

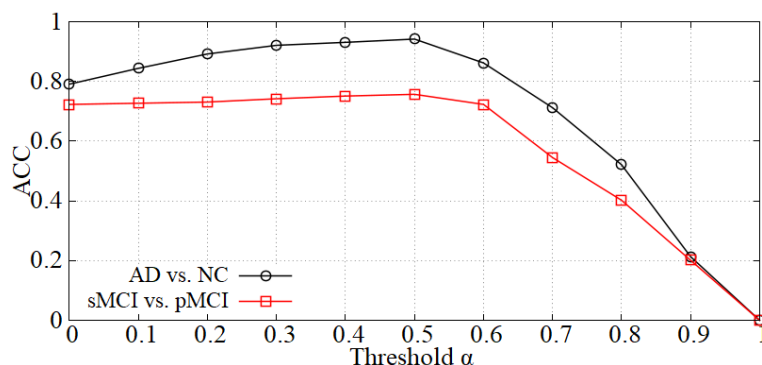


Figure 5. The effect of α on diagnostic accuracy.

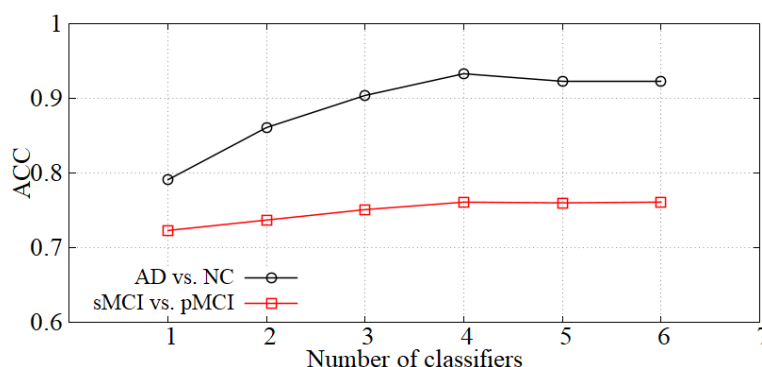


Figure 6. The effect of K on diagnostic accuracy.

4.3.2. Effectiveness of Data Denoising Module

In order to illustrate the effect of the data denoising module, we compare the model with the denoising module and the model without the denoising module in both AD vs. NC and sMCI vs. pMCI tasks. The experimental results are shown in Figure 7. The experimental results show that we can get better diagnostic effects by removing the useless information of 3D MRI boundaries. ACC, SPE, and AUC have been significantly improved in the two diagnostic tasks. In the process of extracting the features of the raw data, the network calculates a large amount of boundary noise which will disturb feature extraction and diagnosis. After we crop the boundary noise of images, the useless voxels are reduced, making the diagnosis effect better in two tasks. Moreover, the training speed is also improved significantly after data noise reduction, as shown in Table 3.

Table 3. Comparison of training time with or without data denoising.

Training Time(s)	AD vs. NC	sMCI vs. pMCI
Without denoising	12,093	13,784
With denoising	6399	7273

In order to see the effect of the data denoising module on the AD diagnosis, we visually show the features extracted from the last fully connected layer in the AD vs. NC and sMCI vs. pMCI diagnosis tasks. As shown in Figure 8, the left sub-figure represents the AD vs. NC diagnosis task, and the right sub-figure represents the sMCI vs. pMCI diagnosis task. The features of a sample can be represented by a point (FC-X, FC-Y) in a two-dimensional space, where (FC-X, FC-Y) is the feature vector from the last fully connected layer of our network. The softmax function is used after the fully connected layer, so the points corresponding to all samples are on the straight line $x + y = 1$. In Figure 8, a blue dot indicates that the label is a positive sample, and an orange dot indicates that the label is a negative sample. In the final prediction, if $y > x$, the test sample is predicted as a positive sample, otherwise it is a

negative sample. Therefore, the dividing line of the network model can be regarded as the straight line $y = x$. From this, it can be seen that the data denoising module improves the network’s ability to distinguish between positive samples and negative samples.

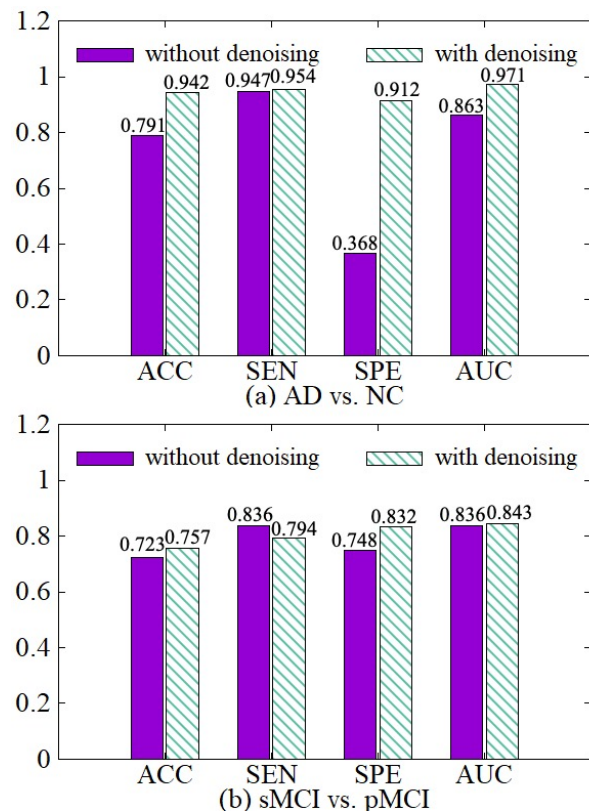


Figure 7. Effectiveness of the data denoising module. (a) shows the performance of AD vs. NC diagnostic task. (b) shows the performance of sMCI vs. pMCI diagnostic task.

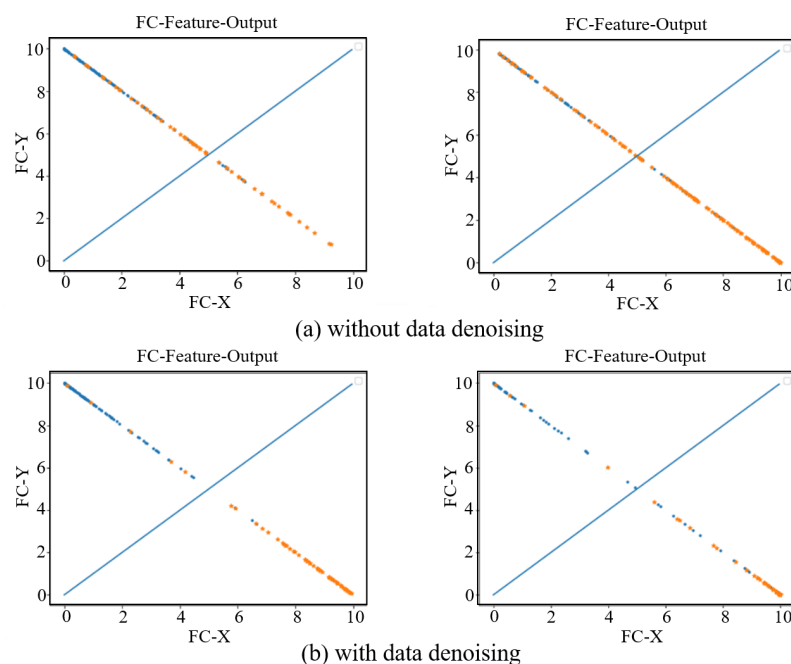


Figure 8. Fully connected layer feature visualization. The left two subfigures represent the AD vs. NC diagnosis task, and the right two subfigures represent the sMCI vs. pMCI diagnosis task. (a) shows the feature visualization without data denoising. (b) shows the feature visualization with data denoising.

4.3.3. Effectiveness of Diagnosis Network Module

We make a comparative experiment which uses a single classifier and our diagnosis network, which incorporates the idea of ensemble learning and contains multiple classifiers. The experimental results are shown in Figure 9. After the diagnosis network module is used, the weights of hard samples are increased, so that the network strengthens the learning of the hard samples. Besides, the diagnosis network learns the fusion weights of various classifiers and the diagnosis results of multiple classifiers are adaptively fused together with their weights in the final fusion diagnosis. So the results of our diagnosis network on the four indicators ACC, SEN, SPE, and AUC are better than the single classifier in both AD vs. NC diagnosis tasks.

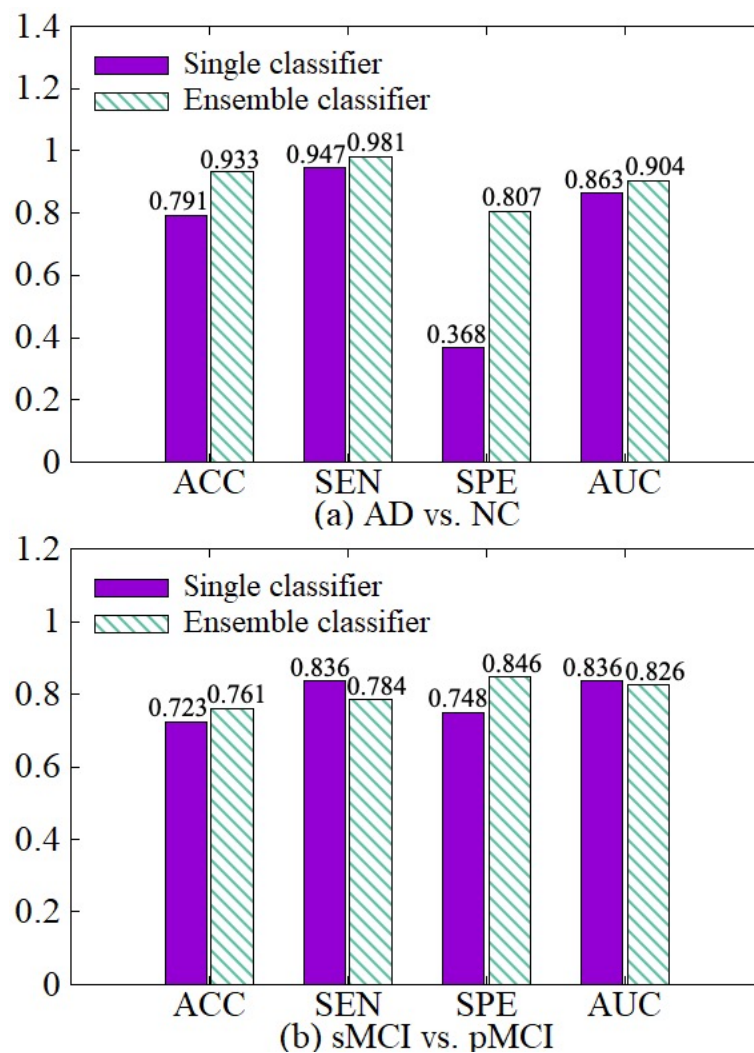


Figure 9. Effectiveness of the diagnosis network module. (a) shows the performance of AD vs. NC diagnostic task. (b) shows the performance of sMCI vs. pMCI diagnostic task.

4.3.4. Effectiveness of the Improved Loss

We compare the diagnosis results based on the traditional cross-entropy loss and our improved loss, as shown in Figure 10. On the basis of the traditional cross-entropy loss, we incorporate the weights of various samples to increase the punishment for hard samples in the training process, so the network can focus on learning hard samples. Besides, the loss function of each classifier is adaptively adjusted with its error. Therefore, the diagnosis results based on our improved loss on the four indicators are better than the traditional cross-entropy loss.

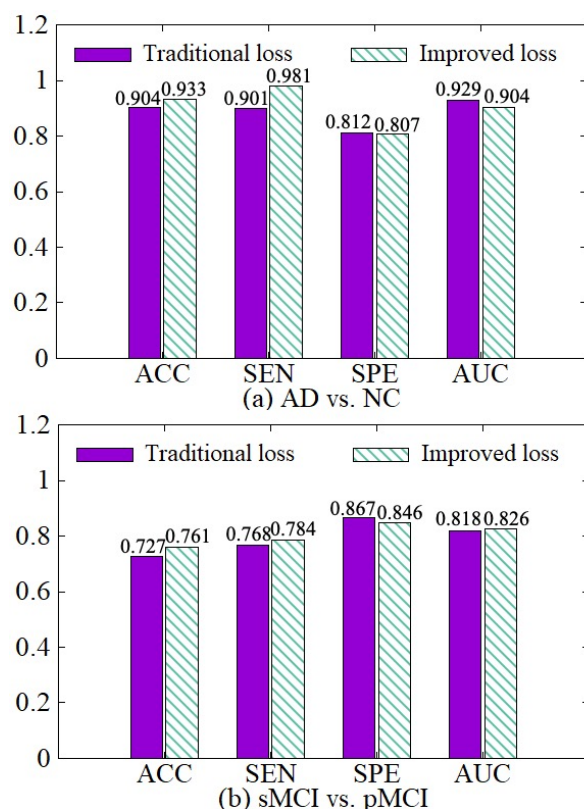


Figure 10. Effectiveness of loss modification. (a) shows the performance of AD vs. NC diagnostic task. (b) shows the performance of sMCI vs. pMCI diagnostic task.

4.3.5. Effectiveness of the Model Including Data Denoising and Diagnosis Network

In order to verify the effectiveness of the model including the data denoising and the diagnosis network, we design an experiment as shown in Figure 11 where various background colors are used to indicate different diagnosis effects, and a darker color indicates a better effect. In Figure 11, W-dd means that the data denoising module is used, N-dd means that the data denoising module is not used. W-dn means that the diagnosis network module is used, and N-dn means that the network is trained with a single classifier. From the experimental results, it can be seen that W-dd and W-dn have achieved good results in both classification tasks. Among them, the two indicators of ACC and SPE have better results. The not very obvious effect of the AUC may be caused by the imbalance of positive and negative samples. In conclusion, the two modules proposed in this paper are effective for the diagnosis of this disease.

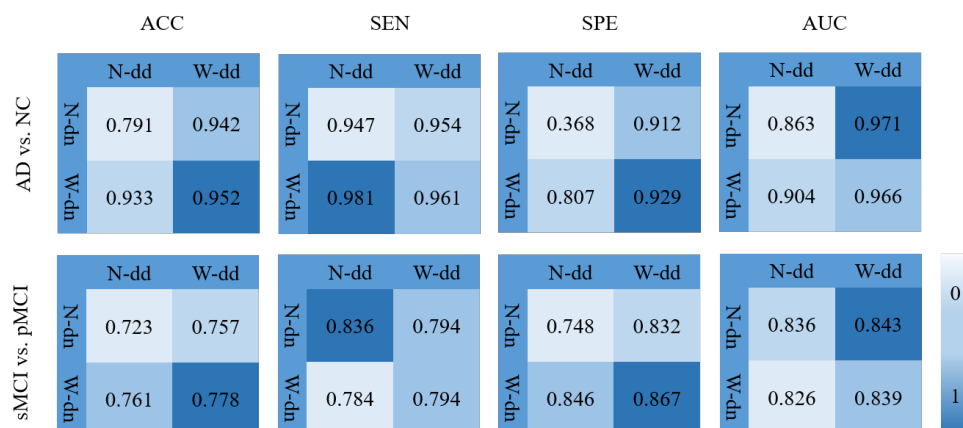


Figure 11. Effectiveness of the model including data denoising and diagnosis network.

4.3.6. Comparisons with Other Methods

To verify the effectiveness of our work, we compare our method with other AD diagnosis methods. It is worth noting that all the baselines are using 3D CNN as the feature extraction basis neural network. In Table 4, a brief description of the state-of-the-art studies using ADNI structural MRI data AD vs. NC diagnostic task and sMCI vs. pMCI diagnostic task is presented. The experimental results show that our method has good performance on both two tasks. Our proposed method outperforms the previous work on both tasks, with an accuracy of 95.2% for AD vs. NC, 77.8% for sMCI vs. pMCI. This fully indicates that our method is effective for the diagnosis of Alzheimer's disease.

Table 4. Comparisons with other methods.

Methods	AD vs. NC				sMCI vs. pMCI			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
Suk et al. [18]	0.92	0.92	0.95	0.97	0.72	0.37	0.91	0.73
Ortiz et al. [20]	0.90	-	-	0.95	-	-	-	-
Liu et al. [16]	0.91	0.88	0.94	0.96	0.77	0.42	0.82	0.78
Karasawa et al. [27]	0.94	-	-	-	-	-	-	-
Cui et al. [11]	0.92	0.91	0.94	0.97	0.75	0.73	0.76	0.80
Feng et al. [30]	0.95	0.98	0.93	0.97	-	-	-	-
Lian et al. [19]	0.90	0.82	0.97	0.95	0.81	0.53	0.85	0.78
Alinsaif et al. [2]	-	-	-	-	0.70	0.60	0.80	-
Ours	0.95	0.96	0.93	0.97	0.78	0.79	0.87	0.84

Moreover, to verify the convergence of the proposed method, as shown in Figure 12, we report the loss function value at each epoch in AD vs. NC diagnosis task and sMCI vs. pMCI diagnosis task. It can be observed that the objective function value decreases in the first 40 epochs and then gradually converges within 50 epochs.

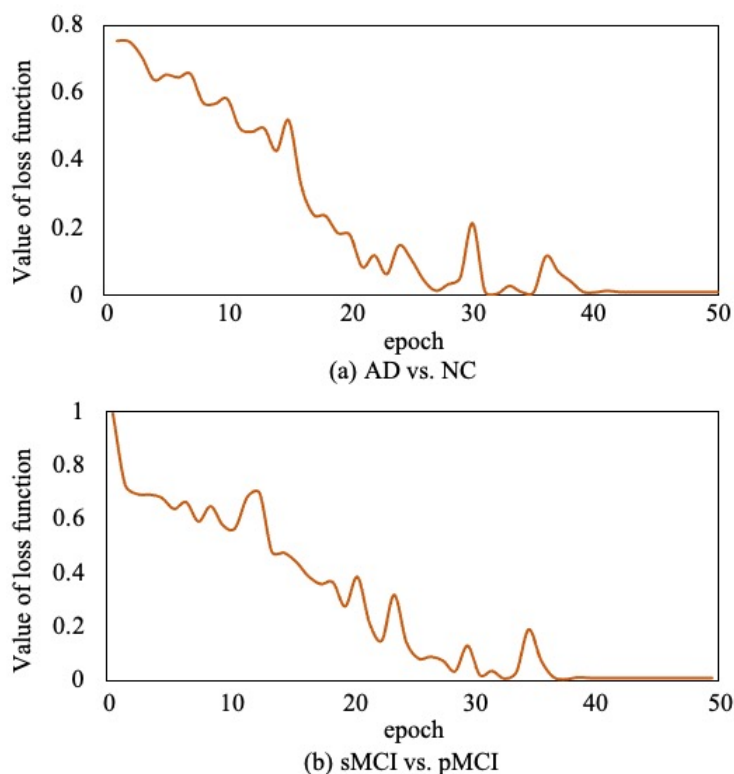


Figure 12. Convergence curves of loss function.

5. Conclusions

In this study, a novel model with ensemble learning classifiers and 3D convolutional neural networks was proposed to diagnose Alzheimer's disease. First, we present an effective data denoising method to reduce data noise so that we can improve the accuracy of diagnosis while speeding up the training process. Second, we propose a diagnosis network with multiple classifiers incorporating the ensemble learning idea to further improve the diagnosis effect, where the limited 3D MRI samples are repeatedly utilized in various classifiers and the weight of each sample in various classifiers can be learned to focus on hard samples. The corresponding loss function is also improved. Next the diagnostic results of various classifiers are adaptively fused based on their separate errors. Finally, extensive experiments have been done on the ADNI dataset to verify the effectiveness of our model. The experimental results indicate that the model proposed in this paper improves the training speed of the neural network and achieves 95.2% accuracy rate in AD vs. NC task and 77.8% accuracy rate in sMCI vs. pMCI task. Compared with several state-of-the-art computer-aided diagnostic methods, our proposed method has demonstrated better or at least comparable classification performance. In this paper, the threshold setting in the data denoising module is fixed for different images by experiments. In the future, we will consider employing dynamic thresholds for various images so as to minimize the loss of effective voxels while reducing noise and further improve diagnosis effect. Once again, thank you very much for your valuable comments and suggestions.

Author Contributions: Conceptualization, P.Z., S.L., J.Q. and Y.T.; methodology, P.Z., S.L. and Y.T.; resources, S.L. and J.Q.; software, P.Z.; writing—original draft, P.Z.; writing—review and editing, S.L., J.Q. and Y.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Science & Technology Development Project of Liaoning Province of China under Grant number 2021JH6/10500127 and Science Research Project of Liaoning Department of Education of China under Grant number LJKZ0008.

Institutional Review Board Statement: A benchmark dataset, Alzheimer's Disease Neuroimaging Initiative (ADNI), was used in our proposed work where the terms of use are declared in the following link: <http://adni.loni.usc.edu/terms-of-use/> (accessed on 16 November 2021).

Informed Consent Statement: A benchmark dataset, Alzheimer's Disease Neuroimaging Initiative (ADNI), which was used in our work, has obtained the informed consent from the participants. More information can be found in the following link: <http://adni.loni.usc.edu/study-design/> (accessed on 16 November 2021).

Data Availability Statement: The dataset used in this study was obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI). More information regarding ADNI can be obtained from the following link: <http://adni.loni.usc.edu/> (accessed on 16 November 2021).

Acknowledgments: This work was supported by Science Research Project of Liaoning Department of Education of China under Grant LJKZ0008 and Science and Technology Development Project of Liaoning Province of China under Grant 2021JH6/10500127.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput.* **1989**, *1*, 541–551. [[CrossRef](#)]
2. Alinsaif, S.; Lang, J. 3D Shearlet-Based Descriptors Combined with Deep Features for the Classification of Alzheimer's Disease Based on MRI Data. *Comput. Biol. Med.* **2021**, *138*, 104879. [[CrossRef](#)] [[PubMed](#)]
3. Baron, J.C.; Chételat, G.; Desgranges, B.; Perchey, G.; Eustache, F. In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease. *Neuroimage* **2001**, *14*, 298–309. [[CrossRef](#)] [[PubMed](#)]
4. Li, S.; Yuan, X.; Pu, F.; Li, D.; Fan, Y.; Wu, L.; Chao, W.; Chen, N.; He, Y.; Han, Y. Abnormal changes of multidimensional surface features using multivariate pattern classification in amnesic mild cognitive impairment patients. *J. Neurosci.* **2014**, *34*, 10541–10553. [[CrossRef](#)] [[PubMed](#)]
5. Zhang, J.; Gao, Y.; Gao, Y.; Munsell, B.C.; Shen, D. Detecting anatomical landmarks for fast Alzheimer's disease diagnosis. *IEEE Trans. Med. Imaging* **2016**, *35*, 2524–2533. [[CrossRef](#)]

6. Liu, M.; Cheng, D.; Wang, K.; Wang, Y. Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis. *Neuroinformatics* **2018**, *16*, 295–308. [[CrossRef](#)] [[PubMed](#)]
7. Zeng, N.; Qiu, H.; Wang, Z.; Liu, W.; Zhang, H.; Li, Y. A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease. *Neurocomputing* **2018**, *320*, 195–202. [[CrossRef](#)]
8. Wang, S.H.; Phillips, P.; Sui, Y.; Liu, B.; Yang, M.; Cheng, H. Classification of Alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling. *J. Med. Syst.* **2018**, *42*, 1–11. [[CrossRef](#)]
9. Cui, R.; Liu, M.; Alzheimer's Disease Neuroimaging Initiative. RNN-based longitudinal analysis for diagnosis of Alzheimer's disease. *Comput. Med. Imaging Graph.* **2019**, *73*, 1–10. [[CrossRef](#)]
10. Ju, R.; Hu, C.; Li, Q. Early diagnosis of Alzheimer's disease based on resting-state brain networks and deep learning. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2017**, *16*, 244–257. [[CrossRef](#)]
11. Cui, R.; Liu, M. Hippocampus Analysis by Combination of 3-D DenseNet and Shapes for Alzheimer's Disease Diagnosis. *IEEE J. Biomed. Health Inform.* **2018**, *23*, 2099–2107. [[CrossRef](#)] [[PubMed](#)]
12. Wang, H.; Shen, Y.; Wang, S.; Xiao, T.; Deng, L.; Wang, X.; Zhao, X. Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer's disease. *Neurocomputing* **2019**, *333*, 145–156. [[CrossRef](#)]
13. Ashburner, J.; Friston, K.J. Voxel-based morphometry—the methods. *Neuroimage* **2000**, *11*, 805–821. [[CrossRef](#)]
14. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
15. Tong, T.; Wolz, R.; Gao, Q.; Hajnal, J.V.; Rueckert, D. Multiple Instance Learning for Classification of Dementia in Brain MRI. *Med. Image Anal.* **2013**, *16*, 599–606.
16. Liu, M.; Zhang, J.; Adeli, E.; Shen, D. Landmark-based deep multi-instance learning for brain disease diagnosis. *Med. Image Anal.* **2018**, *43*, 157–168. [[CrossRef](#)]
17. Pan, Y.; Liu, M.; Lian, C.; Zhou, T.; Xia, Y.; Shen, D. Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 16–20 September 2018; pp. 455–463.
18. Suk, H.I.; Lee, S.W.; Shen, D.; Initiative, A.D.N. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage* **2014**, *101*, 569–582. [[CrossRef](#)] [[PubMed](#)]
19. Lian, C.; Liu, M.; Zhang, J.; Shen, D. Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 880–893. [[CrossRef](#)]
20. Ortiz, A.; Munilla, J.; Gorris, J.M.; Ramirez, J. Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease. *Int. J. Neural Syst.* **2016**, *26*, 1650025. [[CrossRef](#)]
21. Maguire, E.A.; Gadian, D.G.; Johnsrude, I.S.; Good, C.D.; Ashburner, J.; Frackowiak, R.; Frith, C.D.; Cadian, D.C.; Johnsrudet, I.S.; Goodt, C.D. Navigation-related structural change in the hippocampus of taxi drivers. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 4398–4403. [[CrossRef](#)]
22. Wolz, R.; Aljabar, P.; Hajnal, J.V.; L?Tj?Nen, J.; Rueckert, D. Nonlinear dimensionality reduction combining MR imaging with non-imaging information. *Med. Image Anal.* **2012**, *16*, 819–830. [[CrossRef](#)] [[PubMed](#)]
23. Liu, M.; Cheng, D.; Yan, W.; Alzheimer's Disease Neuroimaging Initiative. Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images. *Front. Neuroinform.* **2018**, *12*, 35. [[CrossRef](#)]
24. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
25. Islam, J.; Zhang, Y. Early diagnosis of Alzheimer's disease: A neuroimaging study with deep learning architectures. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 19–21 June 2018; pp. 1881–1883.
26. Maqsood, M.; Nazir, F.; Khan, U.; Aadil, F.; Jamal, H.; Mehmood, I.; Song, O.Y. Transfer learning assisted classification and detection of Alzheimer's disease stages using 3D MRI scans. *Sensors* **2019**, *19*, 2645. [[CrossRef](#)]
27. Karasawa, H.; Liu, C.L.; Ohwada, H. Deep 3d convolutional neural network architectures for alzheimer's disease diagnosis. In Proceedings of the Asian Conference on Intelligent Information and Database Systems, Dong Hoi City, Vietnam, 19–21 March 2018; pp. 287–296.
28. Basaia, S.; Agosta, F.; Wagner, L.; Canu, E.; Magnani, G.; Santangelo, R.; Filippi, M.; Initiative, A.D.N. Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *NeuroImage Clin.* **2019**, *21*, 101645. [[CrossRef](#)] [[PubMed](#)]
29. Duc, N.T.; Ryu, S.; Qureshi, M.N.I.; Choi, M.; Lee, K.H.; Lee, B. 3D-deep learning based automatic diagnosis of Alzheimer's disease with joint MMSE prediction using resting-state fMRI. *Neuroinformatics* **2020**, *18*, 71–86. [[CrossRef](#)] [[PubMed](#)]
30. Feng, C.; Elazab, A.; Yang, P.; Wang, T.; Zhou, F.; Hu, H.; Xiao, X.; Lei, B. Deep learning framework for Alzheimer's disease diagnosis via 3D-CNN and FSBi-LSTM. *IEEE Access* **2019**, *7*, 63605–63618. [[CrossRef](#)]
31. Freund, Y.; Schapire, R.; Abe, N. A short introduction to boosting. *J. Jpn. Soc. Artif. Intell.* **1999**, *14*, 1612.
32. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
33. Nair, V.; Hinton, G.E. Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, Israel, 21–24 June 2010; pp. 807–814.
34. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.

35. Zhou, Y.T.; Chellappa, R. Computation of optical flow using a neural network. In Proceedings of the IEEE 1988 International Conference on Neural Networks, San Diego, CA, USA, 24–27 July 1988; pp. 71–78.
36. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
37. Jack, C.R., Jr.; Bernstein, M.A.; Fox, N.C.; Thompson, P.; Alexander, G.; Harvey, D.; Borowski, B.; Britson, P.J.; Whitwell, J.L.; Ward, C.; et al. The Alzheimer’s disease neuroimaging initiative (ADNI): MRI methods. *J. Magn. Reson. Imaging Off. J. Int. Soc. Magn. Reson. Med.* **2008**, *27*, 685–691. [[CrossRef](#)]
38. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.