

Article

# Multi-View Based Multi-Model Learning for MCI Diagnosis

Ping Cao , Jie Gao  and Zuping Zhang \* 

School of Computer Science and Engineering, Central South University, Changsha 410083, China; ping.cao@csu.edu.cn (P.C.); gaojiexcq@csu.edu.cn (J.G.)

\* Correspondence: zpzhang@csu.edu.cn

Received: 3 March 2020; Accepted: 16 March 2020; Published: 20 March 2020



**Abstract:** Mild cognitive impairment (MCI) is the early stage of Alzheimer’s disease (AD). Automatic diagnosis of MCI by magnetic resonance imaging (MRI) images has been the focus of research in recent years. Furthermore, deep learning models based on 2D view and 3D view have been widely used in the diagnosis of MCI. The deep learning architecture can capture anatomical changes in the brain from MRI scans to extract the underlying features of brain disease. In this paper, we propose a multi-view based multi-model (MVMM) learning framework, which effectively combines the local information of 2D images with the global information of 3D images. First, we select some 2D slices from MRI images and extract the features representing 2D local information. Then, we combine them with the features representing 3D global information learned from 3D images to train the MVMM learning framework. We evaluate our model on the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database. The experimental results show that our proposed model can effectively recognize MCI through MRI images (accuracy of 87.50% for MCI/HC and accuracy of 83.18% for MCI/AD).

**Keywords:** alzheimer’s disease; magnetic resonance imaging; multi-view; CNN

## 1. Introduction

Alzheimer’s disease (AD), a neurodegenerative brain disease caused by multiple factors, is one of the most common chronic diseases in old age [1]. This disease usually causes progressive and disabling impairments of cognitive function, including memory, language, understanding and attention [2]. In 2015, it was estimated that about 47 million people worldwide had AD, and the number is expected to reach 141 million by 2050 [3]. At present, there is no practical method to cure AD [4], so early diagnosis of AD is needed to obtain treatment time. Mild Cognitive Impairment (MCI) is an intermediate state between normal aging and dementia [5], and one study showed that 32% of MCI converted to AD within five years [6]. Therefore, early diagnosis and intervention of Alzheimer’s disease is very important.

In the past few decades, neuroimaging has been widely used to study brain diseases [7–9]. Neuroimaging technology provides anatomical and functional images of the brain, such as Positron Emission Computed Tomography (PECT), Structural Magnetic Resonance Imaging (SMRI), Diffusion Magnetic Resonance Imaging (DMRI), Functional Magnetic Resonance Imaging (FMRI), Electroencephalogram (EEG), and Magnetoencephalography (MEG) [10–12]. Among them, SMRI is often used for the characterization and prediction of AD due to its relatively low cost and good imaging quality. Previous studies have shown that the volume and thickness of the brain are closely related to AD [13], the hippocampus region of AD patients is one third smaller than that of healthy subjects [14], and the medial temporal lobe region is the most effective region of the brain for identifying patients with MCI [15].

In recent years, machine learning and deep learning technologies have demonstrated revolutionary performance in many areas, such as action recognition, machine translation, image segmentation, and speech recognition [16–19]. Machine learning and deep learning have also achieved great success in the fields of medical image analysis and assisted diagnosis of brain diseases [20–23]. Unlike traditional methods based on manual feature extraction, deep learning can learn straightforward and low-level features from medical images, and construct complex high-level features in a hierarchical way [24].

The methods of diagnosing AD through Magnetic Resonance Imaging (MRI) images can be roughly divided into two categories, including 1) methods based on two-dimensional (2D) view, and 2) methods based on three-dimensional (3D) view. In most studies based on 2D view, 2D slices are selected from each subject, and these coronal, sagittal, or axial brain images are considered as a whole to classify. Bi et al. [25] manually selected brain images from three orthogonal panels of MRI data and performed unsupervised learning through PCA-Net. Neffati et al. [26] extracted 2D discrete wavelet transform texture features from coronal slices to classify AD. Jain et al. [27] selected the most informative set of 2D slices and used models trained on natural images to classify medical images through transfer learning. Mishra et al. [28] extracted features through complete local binary pattern from 2D slices in three directions instead of a single direction. These methods based on 2D view have all achieved excellent classification performance.

In addition, there are many methods based on 3D view that can also classify AD well [29–32]. Zhang et al. [33] discriminated between patients and healthy controls by using voxel-based morphometry (VBM) parameters from 3D images. Their results reported that the effect of classification using only 3D VBM parameters was better than the effect of classification using only 2D texture parameters. J. Liu et al. extracted 3D texture features through different divisions of the brain regions of interest (ROI) to construct multiple hierarchical networks [34,35]. Y. Wang et al. [36] diagnosed AD through combining morphometric measures of 3D images and connectome measures of 3D ROI. M. Liu et al. [37] used multivariate statistical tests to find discriminative 3D patch sets from the 3D brain images for the subsequent analysis. Basaia et al. [38] adopted a convolutional neural network to distinguish mild cognitive impairment who will convert to AD (c-MCI) and stable MCI (s-MCI) through 3D MRI images with an accuracy of 75%.

Considering the excellent performance of methods based on 2D view and 3D view, researchers in different fields have combined 2D and 3D views for study. Nanni et al. [39] combined texture features extracted from 2D slices with voxel features from 3D images and used multiple feature selection methods to improve the detection of early AD. In the field of action recognition and human pose estimation, Luvizon et al. [40] used a multi-task learning framework to combine 2D still images with 3D videos to learn features and achieved the latest results at that time. In the field of tumor segmentation, Mlynarski et al. [41] first divided 3D images into 2D images in three directions, and then fed them into the 2D model to generate 2D feature maps. Finally, they used 2D feature maps as the additional inputs of the 3D model to combine 2D and 3D information to achieve better segmentation results. The methods of combining 2D and 3D views in the different fields mentioned above have achieved excellent results. Therefore, we propose a multi-model learning framework based on multiple views to make full use of local and global information in this paper. We combine some 2D MRI images with the entire 3D MRI images, and adopt different deep learning models for different views to better distinguish AD and MCI.

The main contributions are listed as follows:

- (1) We employ entropy to select 2D slices and fuse them to learn 2D local features.
- (2) We propose a combination of 2D and 3D images of MRI to diagnose MCI, rather than a single view.
- (3) We propose a multi-model learning framework that uses different models to train data from different views.

## 2. Materials and Methods

We propose a new method of MCI diagnosis based on multi-view based multi-model (MVMM) framework. The MVMM framework mainly includes a 3D model for extracting global features and a 2D model for extracting local features. The 3D model we use is the Dilated Residual Network (DRN), which adds an Efficient Space Pyramid (ESP) module. The 2D model is the Dual Attention Inception Network (DAIN), which adds a dual attention mechanism to the Inception network. The flow of our MVMM framework is shown in Figure 1. Firstly, the gray matter (GM) images of subjects are divided into whole-brain gray matter images and some selected two-dimensional slices, and then they are respectively input into the corresponding different models. Finally, the local features and global features are concatenated together for integration training.

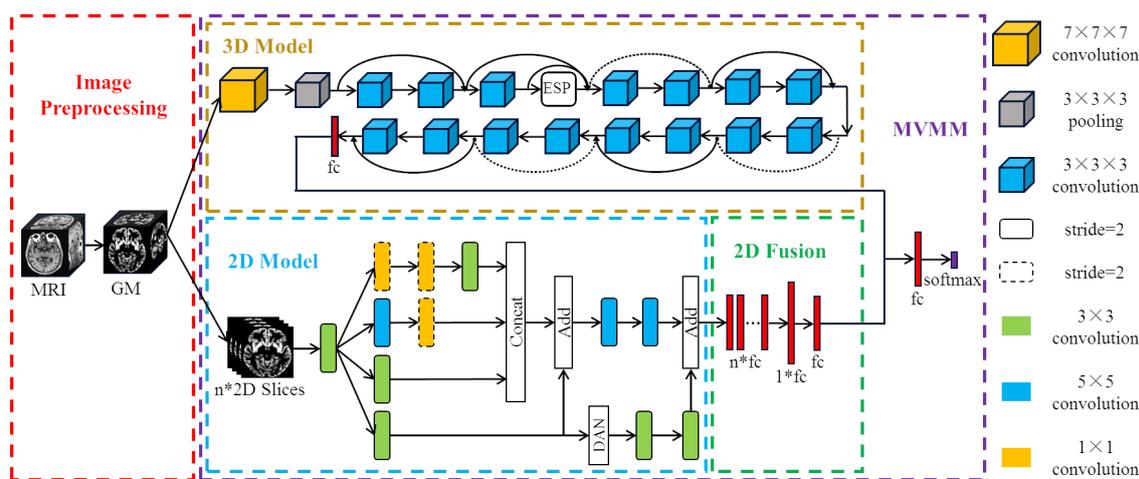


Figure 1. Multi-view based multi-model learning framework.

### 2.1. Data and Pre-Processing

#### 2.1.1. Data Acquisition

In this work, the MRI images we use are obtained from Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu/>) [42]. ADNI started in 2004 under the leadership of Dr. Michael W. Weiner. ADNI is a longitudinal multicenter study with the primary goal of early detection of AD and the use of biomarkers to track disease progression. ADNI has already begun three phases, namely ADNI1 (2004-2009), ADNI2/GO (2010-2016) and ADNI3. At each phase of ADNI, new participants are recruited and agreed to complete various imaging acquisitions and clinical evaluations. Later phases include follow-up scans of some previously scanned subjects and scans of new subjects. In this paper, we select the SMRI data acquired at 1.5 Tesla [43]. The data we use are obtained from 649 subjects, which include 175 scans of AD, 214 scans of healthy controls (HC), and 260 scans of MCI. The demographic and clinical characteristics of all subjects are reported in Table 1.

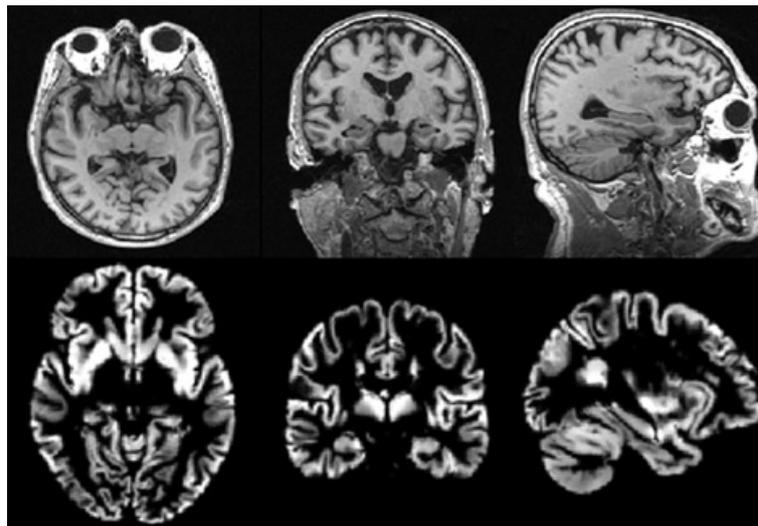
Table 1. Demographic information for 649 subjects.

Class	Subjects	Gender (M/F)	Age	MMSE
AD	175	93/82	75.62 ± 7.38	23.01 ± 2.67
HC	214	115/99	77.09 ± 5.21	29.16 ± 0.98
MCI	260	183/77	75.90 ± 7.37	26.14 ± 3.64

#### 2.1.2. Data Pre-Processing

There is usually much noise in the raw data, so we need to preprocess the MRI data first. In this paper, we use the voxel-based morphological preprocessing method. Specifically, we use the CAT12

toolbox which is an extension to SPM12 [44] to provide computational anatomy. First, we register the MRI images to the standard space through DARTEL (Diffeomorphic Anatomical Registration Through Exponentiated Liealgebra) algorithm [45]. Second, we use the maximum a posteriori and partial volume estimation segmentation techniques [46] to segment the image into gray matter, white matter, and cerebrospinal fluid. Then, the Jacobian determinant is used to modulate the gray matter image nonlinearly. Finally, the gray matter image is spatially smoothed with the 8mm Gaussian smoothing kernel. The size of each gray matter image we get in the standard space is  $121 \times 145 \times 121$ , then we use scikit-image package to resample it to a size of  $96 \times 96 \times 96$ . It is noted that gray matter loss in the medial temporal lobe is characteristic of MCI [47], so we use gray matter images to analyze in this paper. The MRI images before and after preprocessing are shown in Figure 2.



**Figure 2.** Images before and after preprocessing: The first row shows the original images, and the second row shows the final gray matter images.

## 2.2. DRN Model Based on a 3D View

For the whole brain three-dimensional view, we take the preprocessed whole gray matter image directly as input. In order to learn the 3D global information more comprehensively, we use 3D convolutional neural networks to perform global feature extraction related to AD. The 3D gray matter image contains the entire brain, which is very informative. Therefore, how to comprehensively learn useful features is a challenge. Convolutional neural networks have developed rapidly. Since the birth of AlexNet in 2012 [48], the depth of subsequent advanced convolutional neural network models has grown deeper. But as the depth increases, the problem of gradient disappearance during training becomes more serious. In order to avoid the problem of gradient disappearance caused by the network being too deep, the residual network (ResNet) [49] introduces an identity shortcut connection and skips one or more layers directly. Assuming that the layer  $l - 1$  is connected to the layer  $l$ , the output  $x_l$  of the layer  $l$  is:

$$x_l = H_l(x_{l-1}) + x_{l-1} \quad (1)$$

$H(\cdot)$  represents a non-linear transformation function, including batch normalization (BN), rectified linear unit (ReLU), and convolution operation. ResNet reduces the difficulty of training deep networks by adding shortcut connections. For AD-related information, ResNet uses linear activation to obtain identity mapping. In contrast, ResNet uses non-linear activation for redundant information not related to AD. Since the non-linear activation is for redundant information, less useful information is lost. ResNet effectively solves the problem of network degradation caused by too deep depth, so that the model can learn more powerful advanced features.

In this paper, ResNet-18 is used as the basic model to train gray matter images from 3D view to obtain disease characteristics related to AD. However, the traditional ResNet model only uses  $3 \times 3 \times 3$  convolution kernels, and the receptive field of a single convolution kernel can only reach 27. In order to enable the convolution kernel to obtain a larger receptive field, and to allow the model to learn 3D features more comprehensively, we add the dilated convolution [50] on the basis of the ResNet network. Assuming that the size of the convolution kernel is  $k$  and the dilation rate is  $r$ , the receptive field (RF) of the dilated convolution is:

$$RF = [k + (k - 1)(r - 1)]^3, r > 1 \quad (2)$$

When  $r = 1$ , it is ordinary convolution, and the receptive field is  $k^3$ . When  $r = 2$ , for the same  $3 \times 3$  convolution, the dilated convolution not only increases the receptive field (from 9 to 25), but also does not increase the parameters like ordinary convolution (the weight in the dilated part is 0). Although dilated convolutions increase the receptive field, the mesh effect is prone to occur if misused. Therefore, we use an efficient space pyramid (ESP) [51] module to avoid the mesh effect. Specifically, a  $1 \times 1 \times 1$  convolution is performed on the input to obtain a feature map of  $n \times n \times n \times f$ , and then four parallel dilated convolutions are used. Finally, the hierarchical features are merged to obtain the same size feature map. The structure of the ESP module is shown in Figure 3.

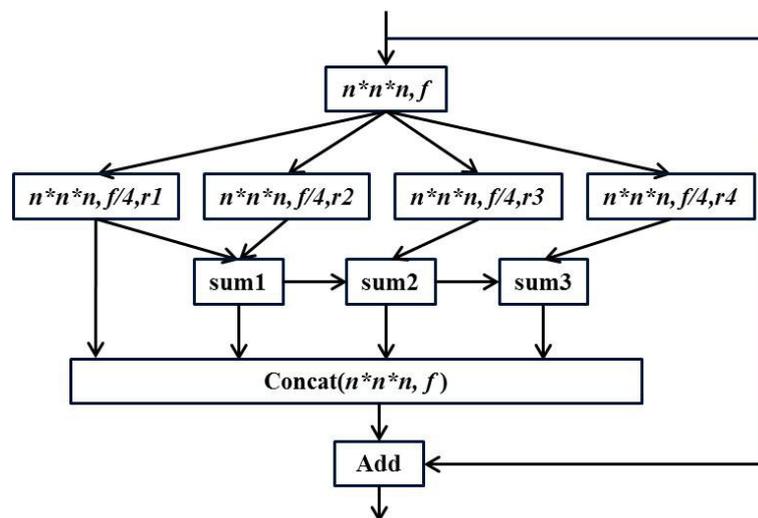


Figure 3. ESP module.

In the figure above, the  $r_1, r_2, r_3, r_4$  are different dilation rates. It can be seen that the residual operation is also used in the ESP module, and the information of different receptive fields is concatenated before the residual operation to ensure the output quality. We replace one layer in the original ResNet-18 network with an ESP module to learn the 3D global features of the brain image more effectively. The overall structure of the final three-dimensional DRN model is shown in Figure 1. The pre-processed whole-brain gray matter image contains much information. In order to effectively learn useful information, we choose 3D deep convolutional neural network for training. We use ResNet-18 with shortcut connections as our base model. At the same time, we also use dilated convolution to expand the receptive field in order to learn AD-related features more comprehensively.

### 2.3. DAIN Model Based on 2D View

For the 2D local view, we select some slices from the preprocessed gray matter image for training. We can select a large number of 2D slices from the 3D gray matter image. How to choose the best training data is very important for the success of the entire method. In this paper, we select 2D images

based on image entropy [52] and extract the most informative slices to train the network. Generally, for a set of  $M$  gray values with probabilities  $p_1, p_2, \dots, p_M$ , the entropy can be calculated as follows:

$$H = - \sum_{i=1}^M p_i \log p_i \quad (3)$$

The higher the entropy, the more information the image contains. The subsequent problem is that MRI images generally contain much noise. Blindly selecting the image with a large amount of information may lead to the selection of some useless images with much noise. The images we study are preprocessed by CAT12. Compared with the original MRI image, the image we use is standardized and smoothed, and the skull of the image is removed. Therefore, we sort the slices in descending order of entropy and select the first 32 images for training to provide robustness according to previous research [53].

After obtaining the selected 2D slices from the 3D MR image, we use them as the inputs of 2D model to learn AD features. Assume that the image set of each subject is  $x_i (x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\})$ , the input of the two-dimensional model can be expressed as:

$$X = x_1, x_2, \dots, x_n \quad (4)$$

Thirty-two slices are selected from each subject through the axial brain image. In order to enhance the representation ability of the model, we use the Inception structure [54] to obtain information on different scales and fuse the features learned by convolution kernels of different sizes. In the 2D model, we pay more attention to the local information as the selected pictures are comparatively informative. How to find the local information that can distinguish MCI from other categories is the focus of our research. The attentional mechanism solves this problem by enabling the model to think globally and focus on more critical local information. We use a dual attention network (DAN) proposed by Fu et al. [55] that combines channel attention and spatial attention. The spatial attention module (PAM) in DAN first performs three  $1 \times 1$  convolutions on the feature map  $A$  to obtain three feature maps  $B, C$ , and  $D$  of the same size ( $h \times w \times c$ ). Moreover, these three feature maps are converted to the size of  $n \times c$  ( $n = h \times w$ ). Then the spatial attention map is calculated from  $B$  and  $C$ . We multiply the spatial attention map by  $D$ , and then multiply the result by the scale coefficient  $\alpha$  (initialized to 0). Finally, the output of PAM is obtained by adding the original feature map  $A$ :

$$Attention_{pam} = \alpha(\text{softmax}(C \cdot B^T))^T \cdot D + A \quad (5)$$

The channel attention module (CAM) in DAN first converts the feature map  $A$  into  $A_r$ , and multiplies  $A_r$  and  $A_r^T$  to obtain the channel attention map with the size of  $c \times c$ . We multiply the channel attention map by  $A_r$  and then multiply it by the scale coefficient  $\beta$  (initialized to 0). The result of the product is converted to the size of  $h \times w \times c$ . Finally, the output of CAM is obtained by adding to the original feature map  $A$ :

$$Attention_{cam} = \beta(A_r \cdot \text{softmax}(A_r^T \cdot A_r)) + A \quad (6)$$

Finally, the results of the two attention modules are added together to form the output of DAN:

$$Attention_{out} = Attention_{pam} + Attention_{cam} \quad (7)$$

We combine DAN with Inception to form the final two-dimensional model. The structure of our proposed DAIN model is shown in Figure 1. The proposed DAIN model first obtains multi-scale AD information through the Inception structure. Then, the dual attention mechanism is used to obtain more significant local information. Finally, the important local information is combined with the fused multi-scale information to obtain the AD features that represent the 2D view.

#### 2.4. Combination of 2D and 3D Views

Before constructing the multi-view model, we first integrate the output of the DAIN Model. Because in the previous DAIN model, we selected 32 slices for each subject and treated each slice as one subject. In this study, we use the above DAIN model as the pre-trained model of the final 2D model. In the final 2D model, the features extracted from each of the 32 images are concatenated together for classification. The final full connection is the MCI feature representing the local information extracted from the 2D slices.

After integrating the 2D features, we combine the features extracted from the 2D model with the features extracted from the 3D model. Specifically, 32 full connections of size 32 are obtained from every MRI image after the DAIN model. Then, through the final 2D model, the 32 full connections of the same subject are concatenated together to obtain the feature of size  $32 \times 32$ . After this layer of full connection, we add another full connection of size 32. That is, the information learned by the same person from the DAIN model is nonlinearly integrated, and then the integrated features are taken as the final two-dimensional AD features. The 2D model and the 3D model are trained separately. Finally, the full connections of the 2D model and the full connections of the 3D model are concatenated together for training. The method of combining 2D with 3D models is our proposed MVMM framework. Our MVMM framework can learn both global information of 3D images and local information of 2D slices.

### 3. Results

#### 3.1. Experimental Setup

We implement the MVMM method through the Tensorflow and compute the model on the NVIDIA TITAN V GPU. The loss function we adopt is binary cross-entropy. We use He normal distribution to initialize the weights of the model. The learning rate of the 2D model is 0.001, while that of the 3D model and MVMM is 0.0001. The evaluation of the proposed method in this paper is based on the following two tasks:

- (1) T1: MCI/HC classification.
- (2) T2: MCI/AD classification.

To evaluate the classification performance, we adopt the 10-fold cross-validation strategy 10 times. We randomly select ten percent of the images from each class as the test set, and the remaining images are further randomly divided into 10 subsets for each category. In the process of cross-validation, each subset is taken as the validation set in turn, and the rest are used as the training set. This process is repeated ten times to get the final result. In this paper, the accuracy (*ACC*), sensitivity (*SEN*) and specificity (*SPE*) are used for evaluation. The three classification performance measures are calculated as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (8)$$

$$SEN = \frac{TP}{TP + FN} \times 100\% \quad (9)$$

$$SPE = \frac{TN}{TN + FP} \times 100\% \quad (10)$$

Among them, *TP* represents true positive, *FP* represents false positive and *FN* represents false negative. In addition, we use the area under the receiver operating characteristic curve (*AUC*) to measure classification performance. The value with the best result for each measure is shown in bold in all tables of experiments.

### 3.2. Experimental Results

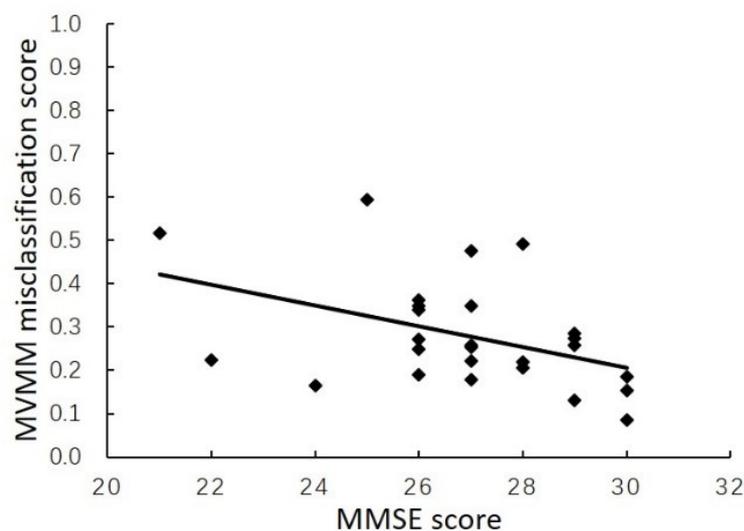
In this section, we present the respective results of the single-view and multi-view models. For the T1 task, it can be observed from Table 2 that the classification performance of MVMM model based on multiple views (87.50% for ACC) is better than the DRN model based on 3D view (81.67% for ACC) and the DAIN model based on 2D view (83.96% for ACC). For the T2 task, it can be observed from Table 3 that although the sensitivity of the MVMM is lower than that of the DAIN model, the ACC, SPE and AUC of the MVMM model (83.18%, 70.56% and 0.8124) are higher than that of DRN model and DAIN model. Therefore, the performance of our MVMM model is the best. In addition, we perform Spearman's Rank Order Correlation analysis to examine the relationship between the misclassification rate of MVMM model and the MMSE score of subject. As shown in Figure 4, the probability of MCI being misclassified as AD is negatively correlated with the MMSE score ( $r = -0.418$ ,  $p = 0.033$ ).

**Table 2.** Classification performance of different views for T1 task.

Model	ACC (%)	SEN (%)	SPE (%)	AUC
DRN	81.67 ± 2.76	84.23 ± 3.53	78.64 ± 5.17	0.8143 ± 0.0311
DAIN	83.96 ± 2.32	87.31 ± 4.22	80.00 ± 3.63	0.8366 ± 0.0320
MVMM	<b>87.50 ± 2.08</b>	<b>89.62 ± 3.86</b>	<b>84.99 ± 3.39</b>	<b>0.8731 ± 0.0219</b>

**Table 3.** Classification performance of different views for T2 task.

Model	ACC (%)	SEN (%)	SPE (%)	AUC
DRN	80.46 ± 3.08	89.62 ± 4.43	67.22 ± 6.23	0.7842 ± 0.0341
DAIN	80.45 ± 2.91	<b>92.31 ± 4.21</b>	63.34 ± 7.11	0.7782 ± 0.0326
MVMM	<b>83.18 ± 2.08</b>	91.92 ± 4.36	<b>70.56 ± 5.04</b>	<b>0.8124 ± 0.0248</b>



**Figure 4.** Relationship between the score of MCI being misclassified as AD and MMSE score.

## 4. Discussion

In order to effectively verify the rationality of the model proposed in this article, we discuss it from the following six aspects: (1) selection of 3D models; (2) selection of 2D models; (3) selection of the number of 2D slices; (4) selection of combination methods; (5) performance on another dataset; (6) comparison with the whole-brain image without segmentation; (7) comparisons with related studies.

#### 4.1. Selection of 3D Models

In this section, we compare the DRN 3D model used in this article with the VGGNet (VN) and DenseNet (DN) models. As shown in Tables 4 and 5, VN and DN model perform poorly in both tasks (T1: 76.04% and 78.33% for ACC; T2: 75.91% and 76.59% for ACC). This is because the 3D MRI images contain a large amount of information, and direct use of the existing network structure cannot get a good effect. The basic model we selected is the ND-DRN model that does not use dilated convolution, which can reduce the redundancy of data by continuously adding shortcut connections in the learning process. Therefore, the ACC obtained by ND-DRN is better than the previous two models (T1: 79.38%; T2: 76.82%). Moreover, for more comprehensive learning, we add dilated convolution to expand the receptive field in the DRN model. Compared with the ND-DRN model, the final DRN model improves the accuracy rate by about 2% to 4%. The ACC, SPE and AUC of our DRN model in T1 and T2 tasks are the highest, while SEN is the second highest. Therefore, from a comprehensive perspective, our DRN model has the best classification performance.

**Table 4.** Classification performance of different 3D models for T1 task.

Model	ACC (%)	SEN (%)	SPE (%)	AUC
VN	76.04 ± 4.08	<b>84.62 ± 5.31</b>	65.90 ± 7.98	0.7526 ± 0.0454
DN	78.33 ± 3.75	79.23 ± 4.45	77.27 ± 5.75	0.7825 ± 0.0381
ND-DRN	79.38 ± 1.96	81.54 ± 4.69	76.82 ± 7.14	0.7918 ± 0.0270
DRN	<b>81.67 ± 2.76</b>	84.23 ± 3.53	<b>78.64 ± 5.17</b>	<b>0.8143 ± 0.0311</b>

**Table 5.** Classification performance of different 3D models for T2 task.

Model	ACC (%)	SEN (%)	SPE (%)	AUC
VN	75.91 ± 3.08	86.54 ± 3.71	60.56 ± 5.67	0.7355 ± 0.0368
DN	76.59 ± 3.05	<b>91.92 ± 3.97</b>	54.44 ± 7.33	0.7318 ± 0.0423
ND-DRN	76.82 ± 2.98	89.23 ± 3.04	58.89 ± 5.96	0.7406 ± 0.0309
DRN	<b>80.46 ± 3.08</b>	89.62 ± 4.43	<b>67.22 ± 6.23</b>	<b>0.7842 ± 0.0341</b>

#### 4.2. Selection of 2D Models

As in Section 4.1, we compare the 2D model used in this article with the classic convolutional network models (AlexNet (AN) and MobileNet (MN)). It can be seen from Tables 6 and 7 that the AUC of the AN and MN models is lower than the NA-DAIN model which does not use attention mechanism. Especially for the T2 task, the SPE of these two models are lower than 50%, which means that they cannot learn the deep local features of the 2D slices. The dual attention mechanism combines spatial and channel attention from a non-local perspective and makes full use of advanced brain features. In this way, more important information can be selected from the 2D data to diagnose MCI better. After adding the attention mechanism, the SPE of the DAIN model reaches 63.34%. At the same time, the ACC and AUC of the DAIN model are also the highest among these models (80.45% and 77.82%).

**Table 6.** Classification performance of different 2D models for T1 task.

Model	ACC (%)	SEN (%)	SPE (%)	AUC
AN	74.37 ± 2.47	80.00 ± 6.84	67.73 ± 7.86	0.7388 ± 0.0243
MN	75.42 ± 4.54	77.69 ± 4.69	72.73 ± 7.34	0.7521 ± 0.0505
NA-DAIN	77.92 ± 3.39	80.00 ± 3.38	75.46 ± 4.09	0.7773 ± 0.0402
DAIN	<b>83.96 ± 2.32</b>	<b>87.31 ± 4.22</b>	<b>80.00 ± 3.63</b>	<b>0.8366 ± 0.0320</b>

**Table 7.** Classification performance of different 2D models for T2 task.

Model	ACC (%)	SEN (%)	SPE (%)	AUC
AN	72.50 ± 2.95	90.38 ± 6.73	46.67 ± 9.15	0.6833 ± 0.0435
MN	69.99 ± 2.22	88.46 ± 5.95	43.33 ± 8.53	0.6593 ± 0.0262
NA-DAIN	75.46 ± 3.49	<b>96.92 ± 4.14</b>	44.44 ± 9.62	0.7068 ± 0.0422
DAIN	<b>80.45 ± 2.91</b>	92.31 ± 4.21	<b>63.34 ± 7.11</b>	<b>0.7782 ± 0.0326</b>

#### 4.3. Selection of the Number of 2D Slices

In the two-dimensional model, we select 32 slices with the largest entropy from each gray matter image. In order to verify the effect of the number of slices on the model performance, we select 8, 16, 24, 32, 40, 48, and 56 slices from each gray matter image for comparison. It can be seen from Tables 8 and 9 that the ACC of the model with eight slices is the worst (T1: 73.13%; T2: 70.68%), which means that too few 2D images cannot represent the brain information of the subject. When the number of slices gradually increases to 32, the ACC improves continuously (T1: from 73.13% to 83.96%; T2: from 70.68% to 80.45%). This indicates that the higher the number of 2D slices, the more disease-related information that can be learned. Nevertheless, it is worth noting that when the number of 2D slices exceeds 32, the ACC no longer improves. It can be inferred that the 32 slices are sufficient to represent the subject's local brain information, and increasing the amount of information will lead to learning local information that is not related to the disease, resulting in a decrease in accuracy.

**Table 8.** Classification performance of 2D models with different numbers of 2D slices for T1 task.

Number	ACC (%)	SEN (%)	SPE (%)	AUC
56	80.79 ± 5.56	86.16 ± 5.75	72.30 ± 3.72	0.7921 ± 0.0485
48	81.88 ± 2.29	84.62 ± 7.49	78.64 ± 8.83	0.8458 ± 0.0290
40	80.00 ± 4.28	83.85 ± 5.65	75.46 ± 5.20	0.7965 ± 0.0459
32	<b>83.96 ± 2.32</b>	<b>87.31 ± 4.22</b>	80.00 ± 3.63	<b>0.8366 ± 0.0320</b>
24	79.17 ± 5.51	77.69 ± 7.25	<b>80.91 ± 7.49</b>	0.7930 ± 0.0562
16	75.83 ± 3.97	80.77 ± 8.60	70.00 ± 6.18	0.7539 ± 0.0365
8	73.13 ± 4.41	78.08 ± 7.70	67.27 ± 8.57	0.7267 ± 0.0444

**Table 9.** Classification performance of 2D models with different numbers of 2D slices for T2 task.

Number	ACC (%)	SEN (%)	SPE (%)	AUC
56	73.64 ± 3.08	91.54 ± 6.61	47.78 ± 6.18	0.6966 ± 0.0281
48	74.32 ± 2.69	91.15 ± 5.96	50.00 ± 8.95	0.7058 ± 0.0312
40	75.78 ± 4.72	92.16 ± 4.50	52.00 ± 9.67	0.7208 ± 0.0525
32	<b>80.45 ± 2.91</b>	<b>92.31 ± 4.21</b>	<b>63.34 ± 7.11</b>	<b>0.7782 ± 0.0326</b>
24	74.09 ± 3.96	80.38 ± 5.66	52.78 ± 8.32	0.7083 ± 0.0460
16	72.27 ± 3.34	85.39 ± 5.09	53.33 ± 9.02	0.6936 ± 0.0373
8	70.68 ± 3.43	86.15 ± 5.57	48.33 ± 5.58	0.6724 ± 0.0440

#### 4.4. Selection of Combination Methods

After obtaining 2D local features and 3D global features, we combine DRN and DAIN for training. Tables 10 and 11 demonstrate that the model has obtained good results, no matter if it is a combination of sum or concatenation. The concatenation fusion method we adopt combines 2D and 3D information from different spaces to achieve higher performance (accuracy of 87.50% for T1 and 83.18% for T2).

**Table 10.** Classification performance of different combination methods for T1 task.

Model	ACC (%)	SEN (%)	SPE (%)	AUC
SUM	86.46 ± 3.12	85.77 ± 4.71	<b>87.27 ± 3.96</b>	0.8652 ± 0.0299
CONCAT	<b>87.50 ± 2.08</b>	<b>89.62 ± 3.86</b>	84.99 ± 3.39	<b>0.8731 ± 0.0219</b>

**Table 11.** Classification performance of different combination methods for T2 task.

Model	ACC (%)	SEN (%)	SPE (%)	AUC
SUM	82.05 ± 2.95	<b>93.08 ± 5.10</b>	66.11 ± 7.03	0.7959 ± 0.0333
CONCAT	<b>83.18 ± 2.08</b>	91.92 ± 4.36	<b>70.56 ± 5.04</b>	<b>0.8124 ± 0.0248</b>

#### 4.5. Performance on Another Dataset

In this section, we validate the generality of our proposed MVMM model on the OASIS (Open Access Series of Imaging Studies) dataset [56]. This dataset contains 3 or 4 individual MRI scans of 416 subjects aged 18 to 96. We use baseline MRI images of 198 subjects aged 60 or older, including 100 AD patients and 98 healthy controls. As shown in Table 12, we can also achieve good performance using the same preprocessing method and MVMM model on OASIS dataset. In other words, the MVMM model we proposed in this paper has potential to be used for wider research.

**Table 12.** Classification performance of MVMM on OASIS dataset.

Dataset	ACC (%)	SEN (%)	SPE (%)	AUC
OASIS	85.65 ± 2.67	85.90 ± 3.83	85.00 ± 3.70	0.8540 ± 0.0291

#### 4.6. Comparison with the Whole-Brain Image without Segmentation

In this paper, the input of our model is the gray matter image after segmentation. In this section, we compare the classification performance using the whole-brain image (without segmentation) and using the gray matter image. It can be seen from Tables 13 and 14 that all classification measures of the model using whole-brain images are lower than the model using gray matter images. This result is because whole-brain images contain more information than gray matter images, and we may need more data to learn the features associated with AD.

**Table 13.** Classification performance of models with different inputs for T1 task.

Input	ACC (%)	SEN (%)	SPE (%)	AUC
whole-brain	83.33 ± 2.46	84.23 ± 5.55	82.27 ± 4.57	0.8325 ± 0.0252
GM	<b>87.50 ± 2.08</b>	<b>89.62 ± 3.86</b>	<b>84.99 ± 3.39</b>	<b>0.8731 ± 0.0219</b>

**Table 14.** Classification performance of models with different inputs for T2 task.

Input	ACC (%)	SEN (%)	SPE (%)	AUC
whole-brain	80.91 ± 2.72	90.82 ± 4.11	66.72 ± 5.71	0.7872 ± 0.0341
GM	<b>83.18 ± 2.08</b>	<b>91.92 ± 4.36</b>	<b>70.56 ± 5.04</b>	<b>0.8124 ± 0.0248</b>

#### 4.7. Comparisons with Related Studies

In this section, we perform experiments to compare the proposed method with other methods. The experimental results are reported in Tables 15 and 16. Islam et al. [57] trained three 3D slices of each participant from the 2D view through ensemble learning. H. Wang et al. and Yuan et al. [58,59] studied the deep features of MRI from a 3D view through the novel convolutional neural network. Although they improved the MRI classification from 2D or 3D views, the verification results using the data in this paper are not as good as our proposed MVMM model. Our proposed MVMM model takes into account the extraction of 2D local information as well as 3D global information, which is more comprehensive than information extracted from a single view. Furthermore, we perform the Student's t test on ACC of different methods to compare the performance. As shown in Tables 15 and 16, our MVMM method is superior to other methods for all classification tasks at  $P < 0.05$ .

**Table 15.** Classification performance of different methods for T1 task.

Model	ACC (%)	SEN (%)	SPE (%)	AUC	P
Islam et al.	80.21 ± 3.26	88.85 ± 3.06	70.00 ± 4.57	0.7943 ± 0.0354	<0.001
H. Wang et al.	83.56 ± 3.38	81.92 ± 6.87	84.13 ± 4.73	0.8304 ± 0.0424	=0.008
Yuan et al.	80.21 ± 2.83	82.31 ± 4.61	77.73 ± 4.49	0.8002 ± 0.0314	<0.001
proposed	<b>87.50 ± 2.08</b>	<b>89.62 ± 3.86</b>	<b>84.99 ± 3.39</b>	<b>0.8731 ± 0.0219</b>	

**Table 16.** Classification performance of different methods for T2 task.

Model	ACC (%)	SEN (%)	SPE (%)	AUC	P
Islam et al.	77.73 ± 3.34	91.16 ± 3.21	58.33 ± 6.48	0.7478 ± 0.0445	<0.001
H. Wang et al.	80.68 ± 1.83	<b>91.92 ± 4.69</b>	64.45 ± 7.53	0.7818 ± 0.0232	=0.014
Yuan et al.	79.32 ± 2.14	89.62 ± 6.89	64.44 ± 6.19	0.7703 ± 0.0358	<0.001
proposed	<b>83.18 ± 2.08</b>	<b>91.92 ± 4.36</b>	<b>70.56 ± 5.04</b>	<b>0.8124 ± 0.0248</b>	

## 5. Conclusions

In summary, we develop a multi-view based multi-model learning framework for the early diagnosis of Alzheimer's disease. First of all, we comprehensively learn global information from the 3D view through residual networks and dilated convolutions. Then we perform the 2D view, which selects the most representative multiple slices through information entropy. Furthermore, we use the Inception network and dual attention mechanism to learn more crucial local information. Finally, we combine the models from 2D and 3D views to train for classification. In this paper, the deep features of MCI are studied both locally and comprehensively by MVMM learning framework. The experimental results show that the proposed method is effective and is expected to be used in the diagnosis of MCI. However, the training of different models in this paper is conducted in parallel, and the concatenation of features is performed at the end. In future research, more complicated combining methods can be considered to make full use of data for MCI diagnosis, such as combining texture features.

**Author Contributions:** P.C. designed and implemented the experiments, analyzed the result, and wrote the paper. J.G. analyzed the result and wrote the paper. Z.Z. analyzed the result and wrote the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Hunan Key Laboratory for internet of Things in Electricity (Grant/Award No. 2019TP1016), the 111 project under Grant No. B18059 and Fundamental Research Funds for the Central Universities of Central South University (Grant No. 2019zzts583).

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

## References

1. Brookmeyer, R.; Johnson, E.; Ziegler-Graham, K.; Arrighi, H.M. Forecasting the global burden of Alzheimer's disease. *Alzheimers Dement.* **2007**, *3*, 186–191. [[CrossRef](#)] [[PubMed](#)]
2. Scheltens, P.; Blennow, K.; Breteler, M.M.B.; de Strooper, B.; Frisoni, G.B.; Salloway, S.; Van der Flier, W.M. Alzheimer's disease. *Lancet Lond. Engl.* **2016**, *388*, 505–517. [[CrossRef](#)]
3. Baumgart, M.; Snyder, H.M.; Carrillo, M.C.; Fazio, S.; Kim, H.; Johns, H. Summary of the evidence on modifiable risk factors for cognitive decline and dementia: A population-based perspective. *Alzheimers Dement.* **2015**, *11*, 718–726. [[CrossRef](#)] [[PubMed](#)]
4. Weiner, M.W.; Veitch, D.P.; Aisen, P.S.; Beckett, L.A.; Cairns, N.J.; Green, R.C.; Harvey, D.; Jack, C.R.; Jagust, W.; Liu, E.; et al. The Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception. *Alzheimers Dement.* **2013**, *9*, e111–e194. [[CrossRef](#)]
5. Facal, D.; Guàrdia-Olmos, J.; Pereiro, A.X.; Lojo-Seoane, C.; Peró, M.; Juncos-Rabadán, O. Using an Overlapping Time Interval Strategy to Study Diagnostic Instability in Mild Cognitive Impairment Subtypes. *Brain Sci.* **2019**, *9*, 242. [[CrossRef](#)]

6. Roberts, R.; Knopman, D.S. Classification and Epidemiology of MCI. *Clin. Geriatr. Med.* **2013**, *29*, 753–772. [[CrossRef](#)]
7. Attallah, O.; Sharkas, M.A.; Gadelkarim, H. Fetal Brain Abnormality Classification from MRI Images of Different Gestational Age. *Brain Sci.* **2019**, *9*, 231. [[CrossRef](#)]
8. Kim, M.; Won, J.H.; Youn, J.; Park, H. Joint-Connectivity-Based Sparse Canonical Correlation Analysis of Imaging Genetics for Detecting Biomarkers of Parkinson’s Disease. *IEEE Trans. Med. Imaging* **2020**, *39*, 23–34. [[CrossRef](#)]
9. Liu, J.; Pan, Y.; Li, M.; Chen, Z.; Tang, L.; Lu, C.; Wang, J. Applications of deep learning to MRI images: A survey. *Big Data Min. Anal.* **2018**, *1*, 1–18.
10. Liu, F.; Wee, C.-Y.; Chen, H.; Shen, D. Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer’s Disease and mild cognitive impairment identification. *NeuroImage* **2014**, *84*, 466–475. [[CrossRef](#)]
11. Kao, C.; Zhang, Y. Magnetic Source Imaging and Infant MEG: Current Trends and Technical Advances. *Brain Sci.* **2019**, *9*, 181. [[CrossRef](#)] [[PubMed](#)]
12. Liu, J.; Pan, Y.; Wu, F.-X.; Wang, J. Enhancing the feature representation of multi-modal MRI data by combining multi-view information for MCI classification. *Neurocomputing* **2020**. [[CrossRef](#)]
13. Chaddad, A.; Desrosiers, C.; Niazi, T. Deep Radiomic Analysis of MRI Related to Alzheimer’s Disease. *IEEE Access* **2018**, *6*, 58213–58221. [[CrossRef](#)]
14. Villain, N.; Desgranges, B.; Viader, F.; de la Sayette, V.; Mézenge, F.; Landeau, B.; Baron, J.-C.; Eustache, F.; Chételat, G. Relationships between Hippocampal Atrophy, White Matter Disruption, and Gray Matter Hypometabolism in Alzheimer’s Disease. *J. Neurosci.* **2008**, *28*, 6174–6181. [[CrossRef](#)]
15. DeCarli, C.; Frisoni, G.B.; Clark, C.M.; Harvey, D.; Grundman, M.; Petersen, R.C.; Thal, L.J.; Jin, S.; Jack, C.R.; Scheltens, P. Qualitative Estimates of Medial Temporal Atrophy as a Predictor of Progression From Mild Cognitive Impairment to Dementia. *Arch. Neurol.* **2007**, *64*, 108–115. [[CrossRef](#)]
16. Chiu, C.-C.; Sainath, T.N.; Wu, Y.; Prabhavalkar, R.; Nguyen, P.; Chen, Z.; Kannan, A.; Weiss, R.J.; Rao, K.; Gonina, E.; et al. State-of-the-Art Speech Recognition with Sequence-to-Sequence Models. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 4774–4778.
17. Luong, M.-T.; Pham, H.; Manning, C.D. Effective approaches to attention-based neural machine translation. *arXiv* **2015**, arXiv:150804025.
18. Piergiovanni, A.J.; Ryoo, M.S. Representation flow for action recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 9945–9953.
19. Zhao, F.; Hu, H.; Chen, Y.; Liang, J.; He, X.; Hou, Y. Accurate Segmentation of Heart Volume in CTA With Landmark-Based Registration and Fully Convolutional Network. *IEEE Access* **2019**, *7*, 57881–57893. [[CrossRef](#)]
20. Sakai, K.; Yamada, K. Machine learning studies on major brain diseases: 5-year trends of 2014–2018. *Jpn. J. Radiol.* **2019**, *37*, 34–72. [[CrossRef](#)]
21. Chen, L.; Bentley, P.; Mori, K.; Misawa, K.; Fujiwara, M.; Rueckert, D. Self-supervised learning for medical image analysis using image context restoration. *Med. Image Anal.* **2019**, *58*, 101539. [[CrossRef](#)]
22. Taheri Gorji, H.; Kaabouch, N. A Deep Learning approach for Diagnosis of Mild Cognitive Impairment Based on MRI Images. *Brain Sci.* **2019**, *9*, 217. [[CrossRef](#)]
23. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [[CrossRef](#)] [[PubMed](#)]
24. Suk, H.-I.; Lee, S.-W.; Shen, D. Deep ensemble learning of sparse regression models for brain disease diagnosis. *Med. Image Anal.* **2017**, *37*, 101–113. [[CrossRef](#)]
25. Bi, X.; Li, S.; Xiao, B.; Li, Y.; Wang, G.; Ma, X. Computer aided Alzheimer’s disease diagnosis by an unsupervised deep learning technology. *Neurocomputing* **2019**. [[CrossRef](#)]
26. Neffati, S.; Abdellafou, K.B.; Jaffel, I.; Taouali, O.; Bouzrara, K. An improved machine learning technique based on downsized KPCA for Alzheimer’s disease classification. *Int. J. Imaging Syst. Technol.* **2019**, *29*, 121–131. [[CrossRef](#)]

27. Jain, R.; Jain, N.; Aggarwal, A.; Hemanth, D.J. Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images. *Cogn. Syst. Res.* **2019**, *57*, 147–159. [[CrossRef](#)]
28. Mishra, S.; Khanna, P. Early detection of Alzheimer's disease using fusion techniques on 3D brain views. In Proceedings of the 2017 Conference on Information and Communication Technology (CICT), Gwalior, India, 3–5 November 2017; pp. 1–5.
29. Kruthika, K.R.; Rajeswari; Maheshappa, H.D. CBIR system using Capsule Networks and 3D CNN for Alzheimer's disease diagnosis. *Inform. Med. Unlocked* **2019**, *14*, 59–68. [[CrossRef](#)]
30. Liu, J.; Wang, J.; Tang, Z.; Hu, B.; Wu, F.-X.; Pan, Y. Improving Alzheimer's Disease Classification by Combining Multiple Measures. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2018**, *15*, 1649–1659. [[CrossRef](#)]
31. Amoroso, N.; La Rocca, M.; Bruno, S.; Maggipinto, T.; Monaco, A.; Bellotti, R.; Tangaro, S. Multiplex Networks for Early Diagnosis of Alzheimer's Disease. *Front. Aging Neurosci.* **2018**, *10*, 365. [[CrossRef](#)]
32. Gupta, Y.; Lee, K.H.; Choi, K.Y.; Lee, J.J.; Kim, B.C.; Kwon, G.R. Early diagnosis of Alzheimer's disease using combined features from voxel-based morphometry and cortical, subcortical, and hippocampus regions of MRI T1 brain images. *PLoS ONE* **2019**, *14*, e0222446. [[CrossRef](#)]
33. Zhang, F.; Tian, S.; Chen, S.; Ma, Y.; Li, X.; Guo, X. Voxel-Based Morphometry: Improving the Diagnosis of Alzheimer's Disease Based on an Extreme Learning Machine Method from the ADNI cohort. *Neuroscience* **2019**, *414*, 273–279. [[CrossRef](#)]
34. Liu, J.; Li, M.; Lan, W.; Wu, F.-X.; Pan, Y.; Wang, J. Classification of Alzheimer's Disease Using Whole Brain Hierarchical Network. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2018**, *15*, 624–632. [[CrossRef](#)] [[PubMed](#)]
35. Liu, J.; Wang, J.; Hu, B.; Wu, F.-X.; Pan, Y. Alzheimer's Disease Classification Based on Individual Hierarchical Networks Constructed With 3-D Texture Features. *IEEE Trans. NanoBiosci.* **2017**, *16*, 428–437. [[CrossRef](#)] [[PubMed](#)]
36. Wang, Y.; Xu, C.; Park, J.-H.; Lee, S.; Stern, Y.; Yoo, S.; Kim, J.H.; Kim, H.S.; Cha, J. Diagnosis and prognosis of Alzheimer's disease using brain morphometry and white matter connectomes. *NeuroImage Clin.* **2019**, *23*, 101859. [[CrossRef](#)] [[PubMed](#)]
37. 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](#)] [[PubMed](#)]
38. Basaia, S.; Agosta, F.; Wagner, L.; Canu, E.; Magnani, G.; Santangelo, R.; Filippi, M. Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *NeuroImage Clin.* **2019**, *21*, 101645. [[CrossRef](#)]
39. Nanni, L.; Brahnam, S.; Salvatore, C.; Castiglioni, I. Texture descriptors and voxels for the early diagnosis of Alzheimer's disease. *Artif. Intell. Med.* **2019**, *97*, 19–26. [[CrossRef](#)]
40. Luvizon, D.C.; Picard, D.; Tabia, H. 2D/3D Pose Estimation and Action Recognition Using Multitask Deep Learning. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5137–5146.
41. Mlynarski, P.; Delingette, H.; Criminisi, A.; Ayache, N. 3D convolutional neural networks for tumor segmentation using long-range 2D context. *Comput. Med. Imaging Graph.* **2019**, *73*, 60–72. [[CrossRef](#)]
42. Jack, C.R.; 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* **2008**, *27*, 685–691. [[CrossRef](#)]
43. Carrillo, M.C.; Bain, L.J.; Frisoni, G.B.; Weiner, M.W. Worldwide Alzheimer's Disease Neuroimaging Initiative. *Alzheimers Dement.* **2012**, *8*, 337–342. [[CrossRef](#)]
44. Friston, K.J.; Holmes, A.P.; Worsley, K.J.; Poline, J.-P.; Frith, C.D.; Frackowiak, R.S.J. Statistical parametric maps in functional imaging: A general linear approach. *Hum. Brain Mapp.* **1994**, *2*, 189–210. [[CrossRef](#)]
45. Ashburner, J. A fast diffeomorphic image registration algorithm. *NeuroImage* **2007**, *38*, 95–113. [[CrossRef](#)]
46. Ashburner, J.; Friston, K.J. Unified segmentation. *NeuroImage* **2005**, *26*, 839–851. [[CrossRef](#)] [[PubMed](#)]
47. Karas, G.B.; Scheltens, P.; Rombouts, S.A.R.B.; Visser, P.J.; van Schijndel, R.A.; Fox, N.C.; Barkhof, F. Global and local gray matter loss in mild cognitive impairment and Alzheimer's disease. *NeuroImage* **2004**, *23*, 708–716. [[CrossRef](#)] [[PubMed](#)]
48. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.

49. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
50. Yu, F.; Koltun, V. Multi-scale context aggregation by dilated convolutions. *arXiv* **2015**, arXiv:151107122.
51. Mehta, S.; Rastegari, M.; Caspi, A.; Shapiro, L.; Hajishirzi, H. Espnet: Efficient spatial pyramid of dilated convolutions for semantic segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 552–568.
52. Studholme, C.; Hill, D.L.G.; Hawkes, D.J. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognit.* **1999**, *32*, 71–86. [[CrossRef](#)]
53. Khan, N.M.; Abraham, N.; Hon, M. Transfer Learning With Intelligent Training Data Selection for Prediction of Alzheimer’s Disease. *IEEE Access* **2019**, *7*, 72726–72735. [[CrossRef](#)]
54. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
55. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual attention network for scene segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 3146–3154.
56. Marcus, D.S.; Wang, T.H.; Parker, J.; Csernansky, J.G.; Morris, J.C.; Buckner, R.L. Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults. *J. Cogn. Neurosci.* **2007**, *19*, 1498–1507. [[CrossRef](#)]
57. Islam, J.; Zhang, Y. Brain MRI analysis for Alzheimer’s disease diagnosis using an ensemble system of deep convolutional neural networks. *Brain Inform.* **2018**, *5*, 2. [[CrossRef](#)]
58. 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](#)]
59. Yuan, L.; Wei, X.; Shen, H.; Zeng, L.-L.; Hu, D. Multi-Center Brain Imaging Classification Using a Novel 3D CNN Approach. *IEEE Access* **2018**, *6*, 49925–49934. [[CrossRef](#)]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).