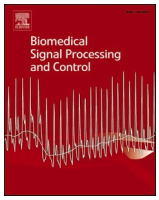
Biomedical Signal Processing and Control 75 (2022) 103565

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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



Multi-modal data Alzheimer's disease detection based on 3D convolution

Zhaokai Kong [a], Mengyi Zhang [a,*], Wenjun Zhu [a], Yang Yi [a], Tian Wang [b], Baochang Zhang [b]

[a] College of Electrical Engineering and Control Science, Nanjing Tech University, Nanjing 210000, Jiangsu, China
[b] Institute of Artificial Intelligence, Beihang University, Beijing 100000, Beijing, China

ARTICLE INFO

Keywords:
Alzheimer's disease detection
MR images
Deep learning
Feature fusion
Positron emission tomography
Convolutional neural networks

ABSTRACT

Multi-modal medical imaging information has been widely used in computer-assisted investigations and diagnoses. A typical example is that the combination of information from multi-modal medical images allows for a more accurate and comprehensive classification and diagnosis of the same Alzheimer's disease (AD) subject. This paper proposes an image fusion method to fuse Magnetic Resonance Images (MRI) with Positron Emission Tomography (PET) images from AD patients. In addition, we use 3D convolutional neural networks to evaluate the effectiveness of our image fusion approach in both dichotomous and multi-classification tasks. The 3D convolution of the fused images is used to extract the information from the features, resulting in a richer multi-modal feature information. Finally, the extracted multi-modal traits are classified and predicted using a fully connected neural network. The experimental results on the Alzheimer's Disease Neuroimaging Initiative (ADNI) public dataset show that the proposed model achieves better results in terms of accuracy, sensitivity and specificity.

1. Introduction

The most frequent kind of senile dementia is Alzheimer's disease. This central nervous system degeneration illness usually has an insidious beginning and progresses slowly. And the corresponding brain failure cannot be reversed. According to data, the aging of the world's population is increasing year by year. Therefore, the number of AD patients with the elderly as the main disease population will inevitably increase [1]. According to a 2018 ADNI research, 50 million individuals globally sustain from dementia, with the figure anticipated to rise to 82 million by 2030, and there will be 152 million dementia patients by 2050, which is three times the number of dementia patients in 2018. At the same time, the report also pointed out that the research on dementia is inefficient. Searching articles showed that there were more than 250,000 articles on dementia and neurodegenerative diseases, and the ratio of articles on tumor and cancer was only about 1:12 [2].

AD is a progressively developing disease, so far the research on AD has not made much breakthrough. Because it is impossible to determine the cause of the illness. As a result, most patients have already reached the advanced stage when the disease is discovered, and missed the critical period of treatment, thus greatly reducing the treatment effect. Therefore, early diagnosis is crucial for the treatment of AD. Mild Cognitive Impairment (MCI) [3] is an intermediate state between AD [4] and health. Patients with MCI seem to be more likely than individuals who have never had the condition to acquire AD, according to studies. And the yearly rate of exchange of MCI to AD might be as high as 10% to 15%. It can be seen that paying more attention to the study of the difference among healthy, MCI, and AD people can greatly contribute to the early diagnosis of AD.

In recent years, non-invasive medical imaging of the brain has been widely used in AD diagnosis [5–7]. This imaging modality is non-invasive to human brain tissues and highly effective based on these advantages [8]. Based on these advantages, non-invasive medical brain imaging has been widely chosen by physicians as one of the most important diagnostic decision aids. Because of the different imaging methods and principles, images of different modalities can highlight different textural structures and regional features of the brain. By learning these features, we can classify and identify patients more effectively, so that the disease can be detected and treated early.While current deep learning approaches are effective in evaluating medical images of AD pathology. Most present research methods slice 3D images and perform 2D convolutional learning, or artificially match 3D brain images with templates to identify specific spatial regions of interest for learning [9]. Nowadays, a three-dimensional convolutional neural network (3D-CNN) [10] is proposed to be applied with magnetic resonance imaging (MRI) to perform binary and ternary disease classification models. The 3D-CNN-Support Vector Machine (SVM) has been shown to give the best results. This approach eliminates the need to

* Corresponding author.
E-mail address: myzhang@njtech.edu.cn (M. Zhang).

https://doi.org/10.1016/j.bspc.2022.103565
Received 9 December 2021; Received in revised form 23 January 2022; Accepted 2 February 2022
Available online 9 February 2022

perform any prior feature extraction manually and is completely independent of the imaging protocol and scanner variability. Simultaneously, many scholars have focused on the studies based on individual modal image data. Using a mix of MRI and PET modal images, we present a model structure based on a 3D CNN to diagnose and categorize AD in this research [11]. Preprocessing is utilized to combine MRI and PET images. The gray matter region of the brain in the MRI image is extracted and integrated into the PET image to achieve image fusion. Finally, a fully connected neural network classifies and outputs this image.

Based on the aforementioned issues, we present a deep learning-based Alzheimer's disease detection strategy. The method employs 3D convolutional neural networks, followed by feature enhancement and multi-modal feature fusion.The following are the paper's main contributions:

- A three-dimensional convolutional network structure is proposed for detecting multi-modal fused features with rich semantic and image details for AD detection. And we add Sparse autoencoder to the network to increase the feature description ability of the network.
- A new image fusion strategy is proposed, where MRI and PET images are fused and fed into the network.
- Through our image fusion method, the features of images can be expressed more intuitively, making our results interpretable.

In this paper, we learn and optimize the proposed method model to achieve higher accuracy and better practical performance in the classification of AD, MCI and NC [12], and to improve the efficiency of timely detection of AD in a comprehensive and accurate way. In turn, further targeted early treatment and effective suppression can be provided to slow down the disease progression and prolong its onset. Experimental results on the classification performance of the ADNI dataset show that the model blocks used to extract features in the proposed model are much smaller in size than the complex 3D convolutional model. In the future, this paper will also optimize and improve the data enhancement, oversampling, and more detailed classification of MCI datasets into sMCI (stable MCI) and pMCI (progressive MCI). The work in this paper can be used as a common way to effectively screen for AD. And our study has value in distinguishing AD and MCI subjects from normal controls. And it can also be valuable for the daily diagnosis of AD in ordinary people.

2. Related work

In the study of AD, some studies sliced PET images along sagittal, coronal and cross-sectional directions, respectively. Then a multi-region image information combination method based on Regions Of Interest (ROI) was used for classification [13]. Since both MRI and PET are three-dimensional images, one study attempted to use MRI images as templates, and then spatially segmented several regions of interest based on the templates after spatial alignment of fluorodeoxyglucose positron emission tomography (FDG-PET) images and MRI image templates for the same subject. Finally, the FDG-PET image region of interest features are extracted and input into a support vector machine model for analysis and prediction [14]. Most of these ROI-based methods are based on prior knowledge in the domain to determine the region of interest. Although these methods are effective, they also have significant limitations.For starters, it might include some human mistakes. Because the signs of Alzheimer's disease are unclear, there is a considerable risk of omission while defining the study area. This will have a significant impact on the outcomes of early diagnosis. Second, most ROI-based illness diagnostic approaches need a considerable quantity of experimental data for training in order to assure the accuracy of prediction, resulting in high time and labor expenses. Besides, complex and diverse preprocessing steps are widely needed to extract medical brain image features in a large number of current studies.

There are many good methods for medical image segmentation in existing deep learning methods. li et al. [15] proposed a dual codec structure of X-shaped network (X-Net). It can be a good alternative to the traditional pure convolutional medical image segmentation network. It can also extract both local and global features to obtain better results. Zhu et al. [16] proposed an image fusion scheme based on image cartoon texture decomposition and sparse representation. The fused cartoon and texture components are fused together according to texture enhancement fusion rules. The experimental results show excellent performance in visual and quantitative evaluation.

Although the ROI-based approach can lead to better actual performance of the model, the native images without feature region extraction or complex preprocessing are also potentially rich in features and performance. Silveira et al. [17] used the Boosting method, which integrates a simple classifier to classify and predict the whole brain PET images. Although this method is plain, its accuracy on AD:NC is 90.97%. This is sufficient to show the inspiration of the method. In addition, Liu et al. [18] also used the whole FDG-PET 3D image sliced in three directions: sagittal, coronal and cross-sectional, and used convolutional neural network to extract features from the slices. The features were then fed into a recurrent neural network for classification, and the classification results of the three directions were finally fused for the final classification prediction. These research methods select only single-modality medical images of the brain for diagnosis and classification of AD. However, multi-modal images provide a large number of useful features for AD pathology, and combining images from different modalities can better represent the features and changes in the human brain for early and more accurate diagnosis of AD.

Although the above methods perform feature extraction for 3D images as a whole, there are still some limitations. That is, the 2D convolution operation is used to extract features of 3D brain images by slicing, which weakens the spatial correlation of 3D images to some extent and potentially enhances the spatial feature loss. For example, Ahemed et al. [19] used an improved 3D convolution model for image segmentation of a 3D image (kidney of African clawed toad) based on the original 2D convolution model. Jose et al. [20] used 3D full convolutional neural network to segment intracranial brain structures, and finally achieved significant results in segmentation of 3D brain MRI images. The experimental results show that there is great potential for plain 3D convolutional operations on the original 3D brain images (MRI images in the experiments). The lightweight 3D convolutional model has good performance compared to even complex models that do not require tedious data preprocessing. However, in the experiments, the authors also consider a potential pitfall: if multiple binary classifications are used instead of a multi-classification task, then it may have the potential problem of "ambiguity zones".

The use of multi-modal image data can show more significant and comprehensive experimental performance and potential value. Different modalities of medical brain images have different imaging principles, which also enable different modal images to highlight the pathological features of different brain diseases. For example, MRI images of brain structures are better able to represent the anatomical structure of the brain [21], which can better represent the anatomical structure and texture of the brain tissue, while PET images are better able to represent the metabolic activity of the brain, which can better detect metabolic abnormalities in pathological regions [22]. Thus, for the same subject, the combined usage of multiple modal images can capture more comprehensive pathological information and features in terms of different characteristics. In addition, Liu et al. proposed a network model based on the combination of MRI image modality and PET image modality for the classification and recognition of AD using multi-modal image data, and achieved satisfactory classification results. However, this method uses a 2D convolutional neural network to operate on 3D images and requires a large number of complex preprocessing steps such as rigid alignment, non-rigid alignment and ROI extraction for each modality image data.

3. Materials and methods

3.1. Datasets

The dataset for this experiment was obtained from the ADNI public dataset. The ADNI is a combination of public and private datasets, which includes multiple sources of biomarker images including MRI and PET images for AD subjects, as well as basic subject information and ratings [23]. ADNI also includes data from multiple institutions at different times, locations and machine models for selection and use [24]. In this paper, we use the public dataset as the basis, according to three categories by retrieving the subjects of pre-processed MRI modal images and FDG-PET images with both T1 weights. MPRAGE-labeled MRI images were chosen because they were thought to be the greatest in terms of quality.

The three classifications of the experimental dataset were AD, MCI, and NC. The number of topics obtained by retrieval was 370. Since the MRI and FDG-PET images of each subject contained several different images at different times, the most recent MRI images of each subject were selected as the MRI modal image data. Also, the first FDG-PET image was selected as the FDG modal image data for this subject by selecting the standardized image. The total number of the experimental data set was 740 3D images after the secondary screening of the data set in the above way.

The number of themes corresponding to the 3 categories was 111 AD themes, 129 MCI themes, and 130 NC themes, respectively. The specific age and gender statistics of the corresponding themes in each category of the dataset are shown in Table 1. It is important to mention that each participant only had one good MRI and one good FDG-PET scan. There are 370 structural MRI modality pictures and 370 FDG-PET modality scans in total. To test the performance of the algorithms in this paper, subsequent experiments were conducted using a 10-fold cross-validation approach.

3.2. Proposed image fusion

To make the multi-modal fusion method more interpretable, we recommend integrating MRI and PET images in the picture area. After that, a single-channel network is used to diagnose the subject using the combined image modality. This approach dramatically decreases the number of design variables when compared to the multi-channel input network with feature fusion. Our suggested AD diagnostic system, which incorporates a heterogeneous image fusion approach, is shown in Fig. 1. Some of the components include image fusion, feature extraction, and classification. We can develop a new medium combining MRI and PET data using our image fusion method.

The suggested multi-modal image fusion approach is capable of combining complimentary information from various modal pictures. As a result, the composite modality describes the information more accurately than a single input picture. As demonstrated in Fig. 2 and Fig. 3, our suggested image fusion approach exclusively recovers the GM area from FDG-PET, which is critical for AD diagnosis, and employs an anatomical mask from an MRI scan. The fusion image mode includes structural MRI data as well as functional PET data. The following are the steps in our image fusion approach in detail.

Table 1
Data set statistics in our experiments.

	AD	MCI	NC
Number of topics/Persons	111	129	130
Age (Average age ± Standard deviation)	76.9 ± 6.7	76.4 ± 6.2	75.7 ± 5.7
Sex (Male/Female)	62/59	78/51	76/54

3.2.1. MRI image processing

Our preprocessing models include FreeSurferare and FSL used in this paper.

The module in FreeSurfer 6.0 is used to conduct skull-stripping on structural MRI images, as illustrated in Fig. 2. The bones and other non-brain material may be removed from the brain size using the watershed segmentation approach, producing in a brain size with significantly fewer distortion and redundant data. The technique retains solely intracranial tissue structure while removing regions of unnecessary anatomical organs, as predicted [25]. As shown in Fig. 2, the FLIRT package in the FSL package is used to affine translate MRI after skull removal to MNI152 space, the global brain map model. FLIRT is an intra- and inter-modal brain object recognition technique that is totally automated, dependable, and exact. The registration procedure aims to minimise translation and revolutions from a normal alignment and erase any spatial variances between individuals in the scanner. The organ separation that follows will be more accurate as a result of this. The input modality for AD classification tasks is this recorded MRI [26]. The GM area is segmented from the MRI image using the FAST module from the FSL software. FAST separates a three-dimensional brain picture into distinct tissue types with taking into account variances in geographic brightness [27]. This sort of tissue is called for bias field inhomogeneities or RF inhomogeneities. The concept is based on a hidden Markov random field model and a related anticipation technique. A slanted adjusted input picture, as well as probability and full volumetric tissue separation, may be obtained from the full automated approach. It is more robust and trustworthy than other limited mixture prototype techniques, which are subject to noise. The segmentation output the gray matter tissue.

3.2.2. FDG-PET image processing

PET image after registration is created by utilizing the FSL FLIRT module to co-register the FDG-PET picture with its corresponding MRI image after registration, as illustrated in Fig. 3. The FDG-PET picture will have the same spatial orientation, image size, and voxel dimensions as the MNI-MRI image. The MNI-PET and MNI-MRI acquired after co-registration are in the same sample space. The anatomical mask created from the gray matter MRI acquired is utilized to cover the whole PET image after registration. A mapping process is used to obtain fusion image, as shown in Fig. 3. So far, we've used FDG-PET scans to determine the anatomical anatomy of GM. However, following MNI152 spatial registration, the projected luminance levels in fusion photos differ significantly from the original PET from radicular and circumferential views, and therefore cannot depict actual metabolic data as well as the origin PET. As shown in Fig. 3, the fusion image is co-registered to the relevant the origin PET image using the FLIRT component to create the fusion image after registration. On the one side, the above enrollment procedure eradicates affine transformation-induced variance and conserves the original PET image's real gray-scale allocation; on the other side, it guarantees that the fusion image after registration has the same dimension of the feature as the origin PET, in other words, the fusion image size of $182 \times 218 \times 182$ is whittled down to the original PET size of $160 \times 160 \times 121$. This drop in quality might potentially conserve processing and memory time.

3.3. Networks

The anatomical backdrop in the direction orthogonal to the 2D plane is totally eliminated in the 2D CNN technique, which processes 3D medical pictures layer by layer. As previously mentioned, using 3D data as a full input can considerably enhance speed, albeit the computational complexity and memory cost rise owing to the high number of factors [28]. As a result, this paper employs 3D CNN, which was created by studying the features of AD classification jobs, as detailed below.

We use a two-step strategy: first, we train a 3D sparse autoencoder to develop convolution layers, and afterwards we create a CNN using the

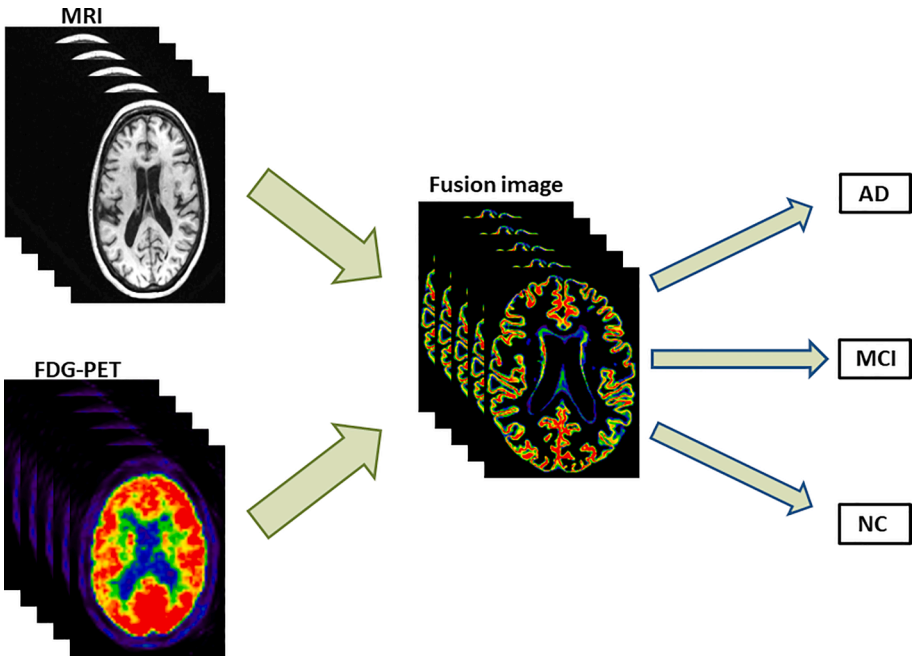


Fig. 1. Proposed AD diagnostic framework with multi-modal image fusion method.

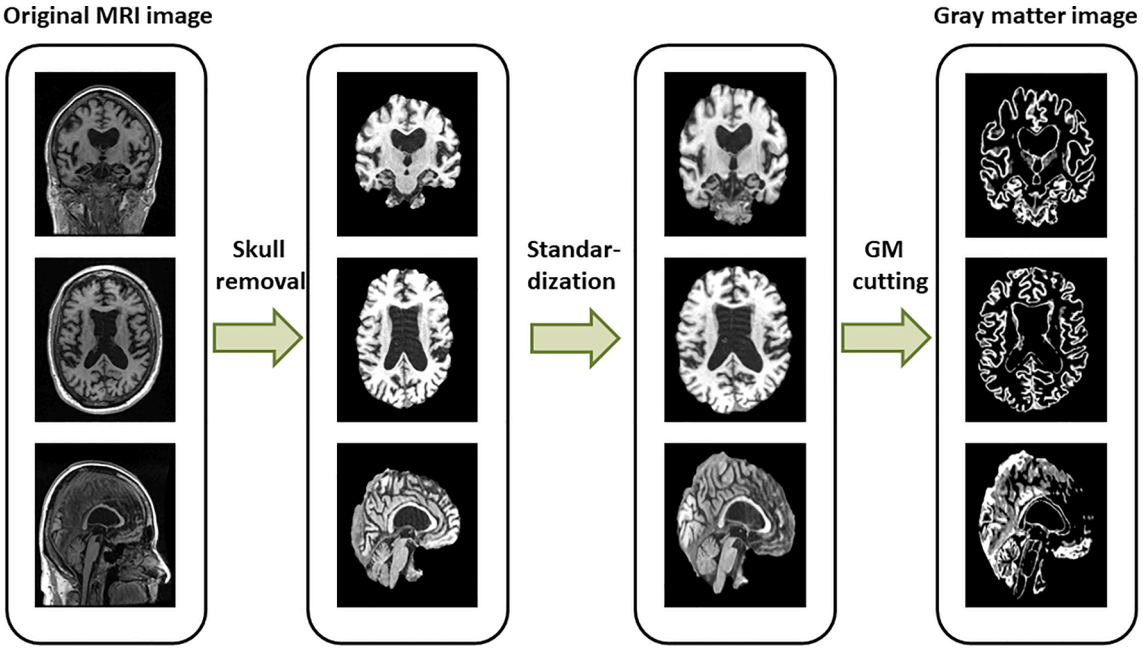


Fig. 2. MRI processing method. Pictures from left to right are the original MRI; the MRI after removing the skull; the registered image; the cropped GM part.

autoencoder's filtration as the first layer.

3.3.1. *Sparse autoencoder*

An autoencoder is a three-layer neural network that uses an image as input to extract features. Sparsity reconstructions may offer a clear explanation of the input information in form of a limited set of features by finding the structure hidden in the data. The autoencoder's source and destination layers have the equal amount of troops, but the hidden layer has more for a scant and complete description. The encoder operator transforms input x to symbol h, while the decoder operator transforms symbol h to output x. In our challenge, we employ 3D regions derived from scanning as the network's input. The decoder operator is in charge of reconstructing the input from h, which is a hidden form. A hidden layer output of 1 means that the node is "active" and a hidden layer output of 0 means that the node is "inactive". Based on this, we introduce the KL dispersion to measure the similarity between the average activation output of a hidden layer node and the sparsity we set.

$$KL = \rho log\frac{\rho}{\widehat{\rho}_j} + (1-\rho) log\frac{1-\rho}{1-\widehat{\rho}_j} \tag{1}$$

Therefore, we can add the KL dispersion as a regular term to the loss function as a way to bound the sparse rows of the whole self-encoder network.

3.3.2. *3D CNN*

The 3D CNN is then trained in the second step. The CNN we utilized

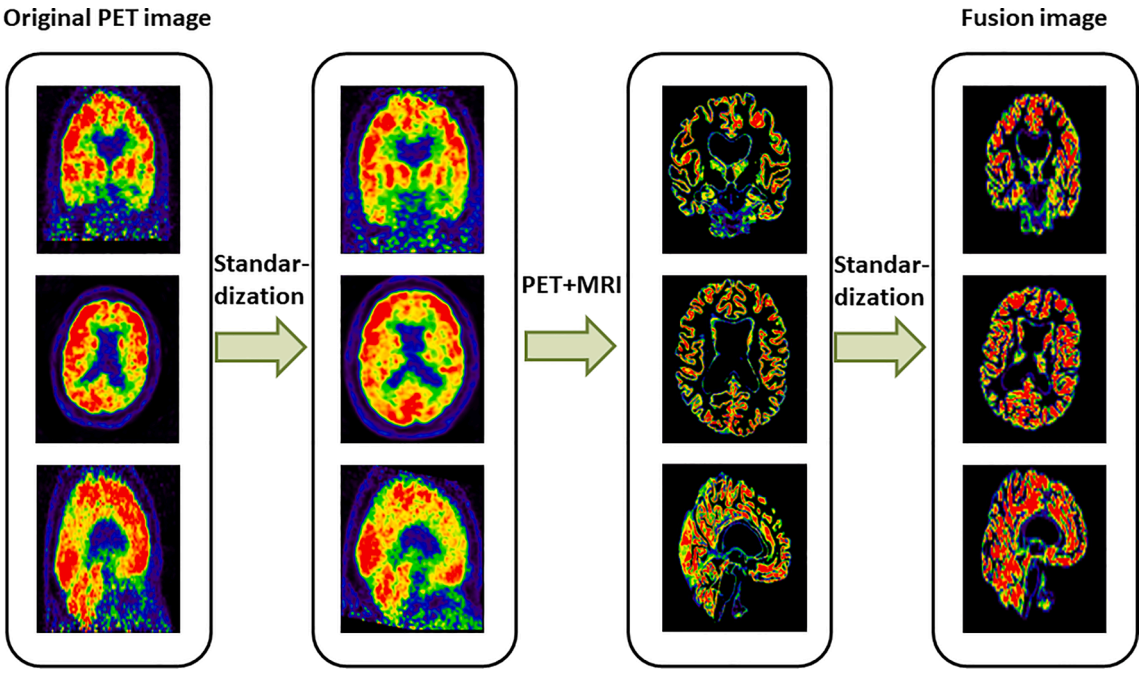


Fig. 3. PET image processing method. The pictures from left to right are the original PET image; the registered PET image; the GM part cropped from the MRI is mapped into the PET image; the final registered fused image.

for this project consists of 1 convolutional layer, 1 pooling layer, 2 linear layers, and finally a log softmax layer. After training the sparse autoencoder, we employ the encoder's weights and biases from the learned architecture in a 3D filter of a 3D convolutional layer of a 1-layer convolutional neural network. Fig. 4 depicts the network's architecture.

4. Experiment and results

4.1. Pre-processing

As inputs to CNN, 3D information with a greater fidelity would need greater computer resources while training stage. To reduce the time it takes to calculate singular information, we manipulate the input data utilizing cropping and sampling techniques. Outside of the brain tissue area, each modality picture contains numerous background areas with pixel values of 0, as illustrated in Fig. 2. We correctly minimize these useless background locations to lower the amount of input information without harming the brain tissue regions. MRI has been trimmed from $182 \times 218 \times 182$ to $176 \times 208 \times 176$ pixels. Furthermore, both PET image and GM-PET image are cropped from $160 \times 160 \times 121$ to $121 \times 145 \times 121$.

4.2. Experimental setup

The main hardware experiment environment is 1 i7-8700 k CPU and 1 GTX3070ti GPU, and the model is built on the basis of Tensorflow framework [29]. The model is trained and tested using a ten-fold cross-validation method after performing a preprocessing process on the raw data. For the binary classification task, after several experiments and parameter adjustments, the final choice of batchsize was set to 3, the number of iterations epoch was set to 200, the learning rate lr was 10^{-7}, and the deactivation rate dropout was 0.5. Such parameter settings enabled the model to achieve the desired performance. In this paper, we also adjusted and tested the parameters for several times, and finally

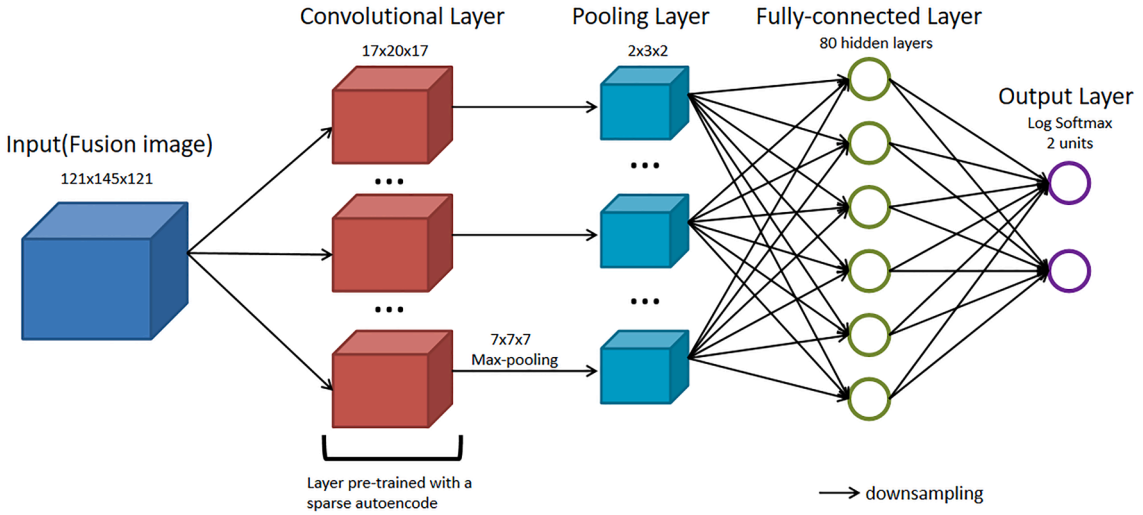


Fig. 4. 3D CNN architecture for AD classification.

determined that the greatest results were obtained while the training data of learning rate was set at 30×10^{-6}, and the other parameters were set the same as the model for the two classification tasks. The model training outputs the training accuracy once per step and the test ACC results once every 10 steps. For the triple classification task, only the test ACC results are output, and the output step interval is set the same as that of the two classification. When the model training is stable, the "test ACC ± standard deviation" is selected as the final model evaluation index.

4.3. Performance

The commonly used performance assessment metrics in AD testing are accuracy, sensitivity, and specificity. Sensitivity is the percentage of true positives, the percentage of positive samples detected out of all positive samples. The higher the sensitivity, the better the detection of positive samples. The greater the specificity, the better the detection of negative samples. The equations for accuracy, sensitivity and specificity are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2)$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

$$Specificity = \frac{TN}{FP + TN} \times 100\% \quad (4)$$

There are four possible predictions for positive and negative cases: positive cases are predicted to be positive (recorded as true positive, TP); positive cases are predicted to be negative (recorded as false negative, TN); negative cases are predicted to be positive (recorded as false positive, FP); and negative cases are predicted to be negative (recorded as true negative, TN).

4.3.1. AD: NC

Table 2 illustrates the findings of unimodal and multimodal modalities with various networks in the categorization of AD: NC. Multi-modality-based techniques like the joint optimization technique and the recommended image compression technique perform better since they effectively fuse MRI and PET data. Our picture fusion approach outperforms other two multi-modal methods in terms of overall indicators. Our image fusion technique achieved the best classification ACC(accuracy) of 93.21±5.0 percent, SPE(specificity) of 95.42±4.5 percent, and SEN(sensitivity) of 91.43±6.0 percent using the 3D CNN. The feature fusion approach had the highest sensitivity (94.44±7.9%) but the lowest accuracy and specificity. In the AD: NC classification test, the actual quality of our image fusion approach was the best.

4.3.2. MCI: NC

Table 3 displays the findings for several treatments in the categorization of MCI: NC using various networks. The proposed picture fusion approach outperformed the competition by a substantial margin. Our picture fusion technique has the greatest sort ACC of 86.52±6.4 percent, SEN of 94.34±6.5 percent, and SPE of 81.64±7.3 percent using the 3D CNN. It also exhibited increases in classification ACC, SEN, and SPE of at least 6.11, 1.25, and 11.62 percent over the feature fusion approach, showing that the suggested image fusion method effectively combines multidimensional data. In general, the suggested image fusion approach outperformed the competition in the MCI: NC classification challenge.

Table 2
Results of different modalities with 3D CNN for AD: NC (UNIT:%).

Network	Modalities	ACC	SEN	SPE
3D CNN	Unimodal MRI	89.80 ± 4.7	86.31 ± 12.0	91.97 ± 5.5
	Unimodal PET	92.10 ± 5.8	89.13 ± 9.7	94.27 ± 4.1
	Proposed image fusion	**93.21 ± 5.0**	**91.43 ± 6.0**	**95.42 ± 4.5**

Table 3
Results of different modalities with 3D CNN for MCl: NC (UNIT:%).

Network	Modalities	ACC	SEN	SPE
3D CNN	Unimodal MRI	79.46 ± 9.4	87.50 ± 16.1	69.15 ± 10.7
	Unimodal PET	72.00 ± 7.8	72.81 ± 10.5	70.56 ± 12.2
	Proposed image fusion	**86.52 ± 6.4**	**94.34 ± 6.5**	**81.64 ± 7.3**

4.3.3. AD: MCI

Table 4 illustrates the findings of two different modalities with various networks in the categorization of AD: MCI. Our image fusion technique for Alzheimer's disease diagnosis used the 3D CNN and achieved the greatest categorization ACC of 85.63±7.8%, SPE of 95.54±6.1 percent, and SEN of 81.21±9.8 percent. Furthermore, the suggested image fusion approach outperformed unimodal methods by at least 6.53, 10.83, and 5.00 percent, respectively, in categorization ACC, SEN, and SPE. In the AD: MCI classification challenge, our technique beat the other methods and had the greatest overall performance.

4.3.4. AD: MCI: CN

The findings of several modalities for the categorization of AD: MCI: NC with the 3D CNN are shown in Table 5. The multi-class task introduces numerous confounding elements since MCI is a transitory condition between AD and NC. Clearly, the AD: MCI: NC classification job is more challenging than the previous binary-categorization problems. In this scenario, our image fusion technique outperformed the unimodal and feature fusion methods on all evaluation indices, but the three-classification task required more power from the unimodal and feature fusion methods. The classification accuracy of the 3D CNN was 87.67±5.1 percent. Our image fusion technique greatly improved the classification accuracy when comparing with the previous methods. Clearly, our image fusion approach outperformed the competition in the multi-class task.

4.3.5. Comparisons with state-of-the-art methods

The suggested image fusion approach was evaluated and contrasted to state-of-the-art multi-modal algorithms for every task-specific categorization. The results are shown in Table 6. For each AD diagnostic task, the findings show that our technique had the best accuracy and outperformed previous multi-modal methods. Although the preprocessing procedures of our image fusion technique are time-consuming, the network parameters are substantially decreased since just the composite picture is given into the categorization networks rather than a set of pictures from several modalities. To put it another way, the proposed picture fusion method's computational difficulty and storage cost are equivalent to that of existing techniques.

4.4. Discussion

For the diagnosis of Alzheimer's disease, we offer a picture fusion approach that successfully combines heterogeneous imaging information from MRI and PET images because multi-modal data can give more

Table 4
Results of different modalities with 3D CNN for AD: MCl (UNIT:%).

Network	Modalities	ACC	SEN	SPE
3D CNN	Unimodal MRI	79.46 ± 9.4	80.32 ± 7.1	69.15 ± 10.7
	Unimodal PET	72.00 ± 7.8	72.81 ± 10.5	70.56 ± 12.2
	Proposed image fusion	**85.63 ± 7.8**	**81.21 ± 9.8**	**95.54 ± 6.1**

Table 5
Results of different modalities with 3D CNN for AD: MCl: NC (UNIT:%).

Network	Modalities	ACC	SEN	SPE
3D CNN	Unimodal MRI	68.00 ± 6.6	–	–
	Unimodal PET	64.32 ± 7.1	–	–
	Proposed image fusion	$\mathbf{87.67 \pm 5.1}$	–	–

Table 6
Comparative performance of our classifiers and competitors.

Approach	Dataset	Accuracy (%)			
		AD: NC	MCI: NC	AD: MCI	AD: MCI: NC
Liu S et al. [30]	MRI + PET	91.4	82.1	–	53.79
Tong T et al. [31]	MRI + PET + CSF + Genetic	91.8	79.5	–	60.2
Li F et al. [32]	MRI + PET	91.4	77.4	70.1	–
Lu D et al. [33]	MRI + PET	84.59	85.96	–	–
Zhu Q et al. [34]	MRI + PET + CSF	88.02	84.14	–	–
Shao W et al. [35]	MRI + PET	92.51	82.53	–	–
Bi XA et al. [36]	fMRI + SNP	81.0	80.0	–	–
Our Method (Image Fusion)	MRI + PET	**93.21**	**86.52**	**85.63**	**87.67**

complete pathology information. Based on the fact that gray matter is the tissues region of most interest in AD diagnostic research, the suggested fusion approach gathers and fuses the gray matter tissue of brain MRI and FDG-PET in the imaging area to generate a fusion modality. As demonstrated in Fig. 2 and Fig. 3, not only does the fusion image maintain the subject's brain structure data from MRI, but it also saves the subject's metabolism data from PET. Furthermore, as compared to techniques based on multi-modal feature learning, our suggested image fusion method handles the problem of heterogeneous features alignment between multi-modal pictures better through its registration operation.

Multimodal medical image fusion is a more intuitive approach compared to existing feature fusion strategies. It integrates relevant and complementary information from multiple input images into a single fused image to facilitate more accurate diagnosis and better treatment. The fused images not only have richer modal features, but also have stronger information representation capabilities.

In addition, the 3D CNN is shown performing 4 Alzheimer's disease categorization tasks, 3 binary and one multi categorization. In addition, the proposed AD diagnosis methodology employs a single network rather than the multiple-input networks utilized in feature matching approaches since our image fusion methodology merges multi-modal image scanning into a single aggregate picture [37]. As a result, the number of CNN parameters may be substantially reduced using our image fusion approach.

To assess the performance of our suggested picture fusion approach, we conducted several tests and studies. According to the classification findings in Tables 2–5, the proposed image fusion method, outperformed unimodal methods because multi-modal methods included more and complementary information. In the challenging three-classification problem, our image fusion method has made great progress compared with previous research. Furthermore, the 3D CNN generated approving results, showing that the image fusion approach had the greatest overall performance and was very adaptable to various classification networks. In addition, when compared to SOTA multi-modal-learning-based methods, our picture fusion method performs better. Although the suggested image fusion approach consistently demonstrated the highest accuracy, its sensitivity and specificity were not always optimum. To address this issue, we will focus even more on WM and CSF tissues in the future, combining their data with the existing gray matter data to give stronger support for Alzheimer's disease diagnosis.

5. Conclusion

For AD diagnosis, we offer an image fusion technique that combines MRI and PET pictures into a composite fusion modality. The fusion modality provides both anatomic and metabolic information about the brain and gently reduces picture noise so that the viewer may focus on the important features. We also proposed improvements to the network. We have added a sparse autoencoder to the network. This allows the network to learn the characteristics that best express the sample in a harsh environment, and can effectively reduce the dimensionality of the sample. The effectiveness of the suggested picture fusion approach was proven by a series of experiments utilizing the 3D CNN. Our proposed image fusion methodology beat the unimodal method by a large margin in terms of experimental outcomes.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by Natural Science Foundation of the Higher Education Institutions of Jiangsu Province of China (21KJB520007); National Natural Science Foundation of China (61803198, 61972016, 62032016, 92067204, 62101245); Beijing Natural Science Foundation (L191007); Fundamental Research Funds for the Central Universities (YWF-20-BJ-J-612).

References

[1] L. Zhan, J. Zhou, Y. Wang, et al., Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer's disease, Front Aging Neurosci. 7 (2015) 48.
[2] P. Christina, World Alzheimer report 2018, Alzheimer's Disease International, London, 2018.
[3] A.M. Sanford, Mild Cognitive Impairment, Clin. Geriatr. Med. 33 (3) (2017) 325–337.
[4] Litjens G, Sanchez, et al. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. Sci. Rep;2016;6:26286.
[5] L. Gao, H. Pan, Q. Li, et al., Brain medical image diagnosis based on corners with importance-values, BMC Bioinform. 18 (1) (2017) 505.
[6] B. Richhariya, M. Tanveer, A.H. Rashid, Diagnosis of Alzheimer's disease using universum support vector machine based recursive feature elimination (USVM-RFE), Biomed. Signal Process. Control 59 (2020), 101903.
[7] R. Hedayati, M. Khedmati, M. Taghipour-Gorjikolaie, Deep feature extraction method based on ensemble of convolutional auto encoders: Application to Alzheimer's disease diagnosis, Biomed. Signal Process. Control 66 (3) (2021), 102397.
[8] L. Gao, H. Pan, X. Xie, et al., Graph modeling and mining methods for brain images, Multimed. Tools Appl. 75 (2016) 9333–9369.
[9] D. Zhang, D. Shen, Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease, Neuroimage 59 (2) (2012) 895–907.
[10] W. Feng, N.V. Halm-Lutterodt, H. Tang, et al., Automated MRI-Based Deep Learning Model for Detection of Alzheimer's Disease Process, Int. J. Neural Syst. 30 (6) (2020) 2050032.
[11] S. Korolev, et al., Residual and plain convolutional neural networks for 3D brain MRI classification, in: IEEE International Symposium on Biomedical Imaging 2017 IEEE, 2017.
[12] Esther E Bron, Marion Smits, Wiesje M van der Flier, Hugo Vrenken, Frederik Barkhof, et al.. Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: the CADDementia challenge. NeuroImage, Elsevier, 2015, 111, pp.562-79.
[13] K.R. Gray, et al., Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease, Neuroimage 60 (1) (2012) 221–229.

[14] K.R. Gray, et al., Regional analysis of FDG-PET for use in the classification of Alzheimer'S Disease, IEEE International Symposium on Biomedical Imaging: From Nano to Macro 2011 (2011) 1082–1085.
[15] Li Y, Wang Z, Yin L, et al. X-Net: a dual encoding-decoding method in medical image segmentation. 2021. doi: 10.1007/s00371-021-02328-7.
[16] Z. Zhu, H. Yin, Y. Chai, et al., A novel multi-modality image fusion method based on image decomposition and sparse representation, Information ences 432 (2018) 516–529.
[17] M. Silveira, J. Marques, Boosting Alzheimer Disease Diagnosis Using PET Images, in: 2010 20th International Conference on Pattern Recognition, 2010, pp. 2556–2559.
[18] Liu, M., D. Cheng, and W. Yan. Classification of Alzheimer's Disease by Combination of Convolutional and Recurrent Neural Networks Using FDG-PET Images. Front. Neuroinform. 12(2018):35.
[19] iek, zgun, et al. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. Springer, Cham (2016).
[20] J. Dolz, C. Desrosiers, I.B. Ayed, 3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study, Neuroimage (2017). S1053811917303324.
[21] M.R. Sabuncu, E. Konukoglu, Clinical Prediction from Structural Brain MRI Scans: A Large-Scale Empirical Study, Neuroinformatics 13 (1) (2015) 31–46.
[22] L. Shen, et al., Semi-supervised manifold learning with affinity regularization for Alzheimer's disease identification using positron emission tomography imaging, International Conference of the IEEE Engineering in Medicine & Biology Society IEEE (2015) 2251–2254.
[23] C.R. Jack Jr, M.A. Bernstein, N.C. Fox, P. Thompson, G. Alexander, D. Harvey, et al., The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods, J. Magn. Reson. Imaging 27 (2008) 685–691.
[24] Liu M, Zhang J, Y ap PT, Shen D. View-aligned hypergraph learning for Alzheimer's disease diagnosis with incomplete multi-modality data. Med Image Anal. (2017) 36:123-34.
[25] A. Bartos, D. Gregus, I. Ibrahim, Brain volumes and their ratios in Alzheimer's disease on magnetic resonance imaging segmented using Freesurfer 6.0, Psychiatry Res. Neuroimaging 287 (2019) 70–74.
[26] V. Fonov, A.C. Evans, K. Botteron, C.R. Almli, R.C. McKinstry, D.L. Collins, Unbiased average age-appropriate atlases for pediatric studies, Neuroimage 54 (2011) 313–327.
[27] M. Jenkinson, P. Bannister, M. Brady, S. Smith, Improved optimization for the robust and accurate linear registration and motion correction of brain images, Neuroimage 17 (2002) 825–841.
[28] F. Milletari, S.A. Ahmadi, C. Kroll, A. Plate, V. Rozanski, J. Maiostre, et al., Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound, Comput. Vis. Image Und. 164 (2017) 92–102.
[29] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, et al., TensorFlow: A system for large-scale machine learning, in: 12th USENIX Symposium on Operating Systems Design and Implementation, USENIX Association, Savannah, GA, 2016, pp. 265–283.
[30] S. Liu, S. Liu, W. Cai, H. Che, S. Pujol, R. Kikinis, et al., Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease, IEEE Trans. Biomed. Eng. 62 (2015) 1132–1140.
[31] T. Tong, K. Gray, Q. Gao, L. Chen, D. Rueckert, Multi-modal classification of Alzheimer's disease using nonlinear graph fusion, Pattern Recogn. 63 (2017) 171–181.
[32] F. Li, L. Tran, K.H. Thung, S. Ji, D. Shen, Li J.A robust deep model for improved classification of AD/MCI patients, IEEE J. Biomed. Health Inform. 19 (2015) 1610–1616.
[33] D. Lu, K. Popuri, G.W. Ding, R. Balachandar, M.F. Beg, M. Weiner, et al., Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images, Sci. Rep. 8 (2018) 1–13.
[34] Q. Zhu, N. Yuan, J. Huang, X. Hao, D. Zhang, Multi-modal AD classification via self-paced latent correlation analysis, Neurocomputing 355 (2019) 143–154.
[35] W. Shao, Y. Peng, C. Zu, M. Wang, D. Zhang, Hypergraph based multi-task feature selection for multimodal classification of Alzheimer's disease, Comput. Med. Imaging Graph. 80 (2020), 101663.
[36] X.A. Bi, X. Hu, H. Wu, Y. Wang, Multimodal data analysis of Alzheimer's disease based on clustering evolutionary random forest, IEEE J. Biomed. Health Inform. 24 (2020) 2973–2983.
[37] Zhang M, Kong Z, Zhu W, et al. Pulmonary nodule detection based on 3D feature pyramid network with incorporated squeeze-and-excitation-attention mechanism. Concurrency and Computation: Practice and Experience, 2021: e6237.