(Received 22 January 2025, Accepted 26 February 2025, Scheduled 12 March 2025)

Diagnosing Alzheimer's Disease Using Multimodal Fusion Neural Network and Weight Optimization for 3-Axial-Plane Patches of MRI and PET

Weiming Lin*, Shumin Feng, Hongfeng Wu, and Jinlin Ma

School of Opto-Electronic and Communication Engineering, Xiamen University of Technology, Xiamen 361024, China

ABSTRACT: Alzheimer's disease (AD) is a brain degenerative disease, so the Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) of cerebral images are effective data in detecting the onset of the disease. In this work, a framework consisting of a cross attention multimodal fusion deep neural network and a patches weight optimization strategy is proposed. First, multiple points are randomly selected from the region of interest (ROI), and multiple 3-axial-plane patches are extracted centered on these points. Then, the patches from MRI and PET are fused using a fusion neural network to output diagnostic information for each patch. Finally, a weight is set for each patch, and a particle swarm optimization algorithm is used to find the optimal weights for multiple patches; the diagnostic information from multiple patches is merged using these weights; and the final diagnostic results are output. The experiments on ADNI dataset show that this method has an accuracy of 94.03% in diagnosing AD and outperforms other methods of fusing MRI and PET data, which demonstrates the promising performance of this method.

1. INTRODUCTION

Alzheimer's disease is a neurodegenerative disorder that usually occurs in the elderly population. The patient's brain is affected by amyloid deposits, and nerve cells are destroyed, thus affecting the patient's brain function. The patient may suffer from memory loss, language disorders, disorientation, etc. Ultimately, it may affect the body function and thus lead to death [1]. The onset of Alzheimer's disease is a gradual process, patients may experience several years from the happening of brain lesions to the noticeable symptoms. Detecting brain lesions in patients through cerebral medical imaging can therefore detect the onset of the disease in a timely manner.

Since Alzheimer's disease causes damage to nerve cells and disruption of connectivity, especially pathological changes in the hippocampus [2], structural Magnetic Resonance Imaging (MRI) can be used to detect atrophy of the brain in Alzheimer's patients [3, 4], especially in the hippocampus. Many studies have been done using different machine learning techniques to identify AD with MRI images. A hybrid MRI meta-ROI for AD detection was introduced by Zheng [5], and the best-performing cortical and volumetric ROIs were used as input nodes to the artificial neural network for AD classification after the detectability of AD in certain cortical and volumetric brain regions had been investigated. Al Shehri [6] proposed a deep learning-based method for the MRI diagnosis and categorization of Alzheimer's disease that makes use of DenseNet-169 and ResNet-50 architectures. Ghosh et al. [7] trained a MobileNet in a federated manner to identify Alzheimer's disease from brain MRI data, and the proposed model uses coronal plane images after an orientation selection procedure. Liu et al. [8] applied statistical and unsupervised learning techniques, and used a small number of labeled structural MRI images, to distinguish between scans from normal control (NC) and those with AD.

On the other hand, the Positron Emission Tomography (PET) scan, such as Fluorodeoxyglucose PET (FDG-PET), is also widely used in the diagnosis and research of Alzheimer's disease [9], since it can objectively evaluates neuronal activity in the brain. As an objective biomarker of neuronal damage, like MRI, FDG-PET has been shown to be an important research standard that can play an important role in both the diagnosis of AD and the determination of the severity of the disease. Moreover, FDG-PET is useful for the early diagnosis of AD because it can show the characteristic patterns of Alzheimer's disease neurodegeneration in people with mild cognitive impairment (MCI) earlier than MRI. For example, De Santi et al. [10] developed a convolutional neural network (CNN) that can classify volumetric 18F-FDG PET images into multiple AD classes. They also tested two distinct post hoc explanation methods that are developed in the Explainable Artificial Intelligence field. Bi et al. [11] investigated the quantitative cerebral blood flow and glucose metabolism in AD patients as determined by FDG-PET diagnostic values. Frings et al. [12] used deformation-based morphometry to test whether cerebral atrophy could be imaged with FDG-PET, and the findings revealed abnormalities in AD, primarily in the bilateral hippocampus. In order to detect AD-related metabolic patterns in neurodegenerative subjects with FDG-PET, Beheshti et al. [13] employed a machine learning-based Alzheimer's disease designation framework. They then generated subject scores using five distinct machine-learning algorithms for AD diagnosis.

* Corresponding author: Weiming Lin (linwm@xmut.edu.cn).

Although MRI and PET can be used individually for the diagnosis and research of AD, data from different modalities reflect the progression of the AD condition from different aspects. Therefore, fusion of MRI and PET data can further improve the diagnostic accuracy of AD. Many researches have been conducted on multimodal data fusion for the diagnosis of AD. For example, using multimodal neuroimaging data, Hojjati and Babajani-Feremi [14] predicted the neuropsychological scores and looked into the nonlinear progression trend of the cognitive declines. The multimodal approach improved the clinical dementia score prediction performance in the converting and stable stages of MCI and AD compared to a single modal approach. In order to learn the interaction between modalities for AD diagnosis with multi-modal data, Zhang et al. [15] proposed a novel multi-modal cross-attention AD diagnosis framework. This framework makes use of cross-modal attention to reinforce relationships between modalities.

Deep neural networks have received extensive attention and research in the automatic detection of medical images [16, 17], but for brain MRI and PET images are 3D data, and the 3D neural network model is hard to be implemented with a complex architecture, which are constrained by the need for high computing capability and graphics memory. In this study, we used multiple 3-axial-plane patches extraction from 3D neuroimaging to reduce the demand on device performance while still serving the purpose of data augmentation. A neural network architecture for multimodal data fusion was then built, which consisted of a two-way ResNet-18 architecture [18] and a content-guided attention (CGA) [19] module to fuse the patches from MRI and PET and output AD diagnostic information for each pair of MRI and PET patches. For the center points of patches at different points in the region of interest (ROI) of the neuroimaging, their contribution to the final diagnosis would be different. Therefore, we assigned different weights to the outputs of patches before merging. In order to find the optimal weights, particle swarm optimization (PSO) algorithm was employed. After several PSO iterations, the optimal weights were found, and all the patches diagnostic information was merged by weights to generate the final AD diagnosis.

The main contributions of this work are as follows:

1. A deep neural network architecture for the multimodal fusion of MRI and PET 3-axial-plane patches was designed to effectively fuse multimodal data and improve AD diagnosis.

2. PSO algorithm was used to optimize the weights of multiple 3-axial-plane patches to improve the combined patches diagnosis.

2. MATERIALS AND METHODS

2.1. MRI and PET Samples and Image Preprocess

The ADNI database, known as the Alzheimer's Disease Neuroimaging Initiative (ADNI), is a large, multicenter, long-term, publicly available database. The goal of ADNI is to measure the progression of AD through quantitative neuroimaging techniques (e.g., MRI and PET) and other biomarkers, as well as through clinical and neuropsychological measures. We downloaded 2031 pairs of MRI and PET images from 945 participants, which means that some images were obtained from different follow-up examination times of the same participant. The ages and clinical dementia scores, including mini-mental state examination (MMSE) and clinical dementia rating (CDR), of these samples are listed in Table 1. These MRI and PET images needed to be preprocessed because these images differed in size, head position, and orientation. The preprocessing includes skull stripping and rigid registration. The MRI images were firstly subjected to N4 bias field correction using the ANTs tool, then skull stripping using the FSL software, rigorous registration to the MNI152 template using the IRTK tool, and finally created a $182 \times 218 \times 182$ sized 3D brain image. The PET images were first registered with the raw MRI image, skull stripped using the MRI skull stripping mask, then aligned with the registered MRI. The preprocession of images are shown in Fig. 1. After registration, all images have the same size and orientation. To avoid data variability introduced by inconsistencies in image brightness, after the preprocessing process, all images were normalized by subtracting the mean and dividing by the standard deviation.

TABLE 1. Ages and clinical dementia scores of samples.

Group	Samples	Age	MMSE	CDR
NC	648	75.9 ± 5.8	29.0 ± 1.1	0.05 ± 0.5
MCI	741	74.4 ± 7.5	27.2 ± 2.1	1.71 ± 1.0
AD	642	75.9 ± 7.3	22.0 ± 3.9	5.31 ± 2.5

After preprocessing, all images were utilized for 3-axial-plane patch extraction. The bounding boxes of the left and right hippocampi were treated as two ROI regions, each with a size of $30 \times 40 \times 45$ according to the size and orientation of the hippocampus. Then, 64 points in each ROI region were randomly selected as extraction locations for the 3-axial-plane patches, so a total of 128 random points were generated. To avoid duplicate points, during the generation process, each new point is compared with previous points, and if the minimum distance from these previous points is less than 2 voxels, the point is discarded, and new points continue to be generated until the number of points meets the requirement. Finally, after all 128 random points were generated, they were utilized for all the images for patch extraction. Three 80×80 size patches were extracted from coronal, transverse, and sagittal planes centered on each point, and these three patches were combined into a 3-channel image. The reason for choosing the patch size of 80×80 is that although the center points of these patches are distributed inside the ROI region, outside the ROI region there is also some information that is useful for diagnosing AD. So the larger the size of the patch is, the more information it could contain. However, if the size is too large, the patch would go beyond the edge of the image, so the size of 80×80 was chosen, which is large and will not go beyond the edge of the image. After 128 3-axial-plane patches were extracted from each image, these patches were fed into the neural network model for the next processing. The extraction of 3-axial-plane patches is demonstrated in Fig. 2.

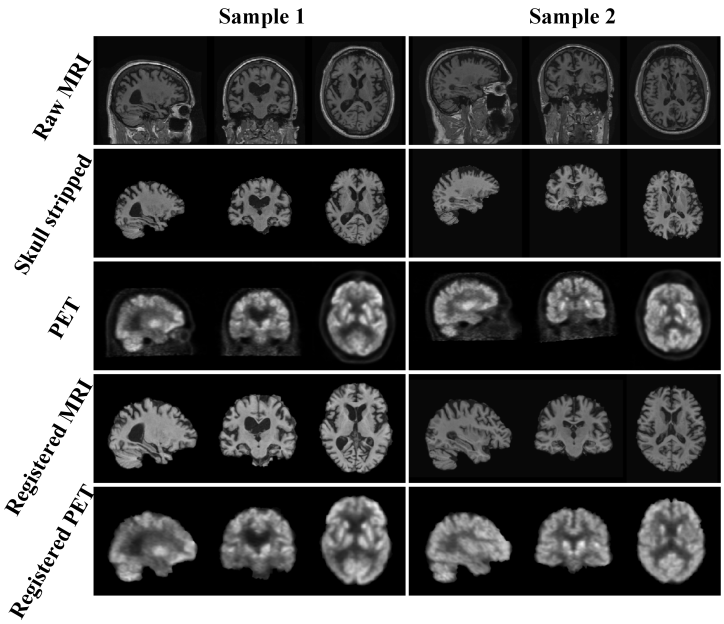

FIGURE 1. The preprocession of MRI and PET scans (2 samples).

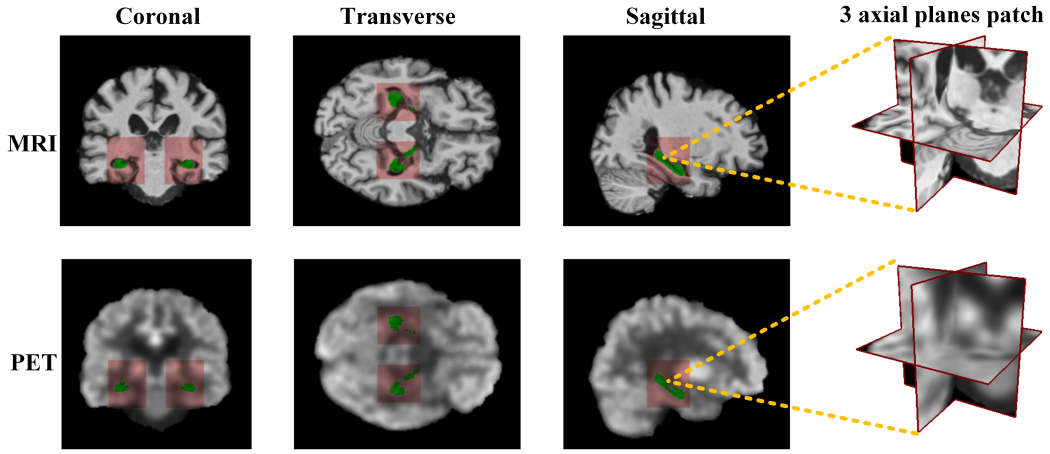

FIGURE 2. The extraction of 3-axial-plane patches. The green regions are left and right hippocampal, and the red regions are ROIs.

2.2. Multimodal Patches Fusing Deep Neural Network Architecture

After 3-axial-plane patch extraction, each sample will get 128 pairs of MRI/PET patches. As shown in Fig. 3, each pair of patches will be processed by a multimodal fusion neural network, and then 128 results will be obtained. These pieces of patch diagnostic information are weighted and merged to get the final diagnostic result. PSO algorithm will be used to find the optimized weighting coefficients to improve the merging effect.

The multimodal fusion neural network mainly consists of two ResNet-18 models, which process MRI and PET patches individually. The structure of ResNet-18 can be divided into three parts: A 7×7 convolutional layer at the beginning of the network for extracting the low-level features of the input image; the main body of the network consists of multiple ResNet layers, each of which contains two residuals blocks, and each residual block has 2 convolutional layers; a fully-connected layer at the end of the network, which is used for mapping the features into the classification space and outputting the final classification results. It should be noted that, starting with ResNet-layer2, the stride of the first convolutional layer of the first residual block is 2, so the size of the feature map output from each ResNet layer will be one-half of the size of the input.

To fuse information of MRI and PET features efficiently, CGA was utilized [19], which could enhance the interrelationship between modalities through an attentional mechanism to optimize the information complementarity between these two modalities. In the two-way ResNet-18 structure, the MRI and PET feature maps output from ResNet-layer2 are taken as inputs to the CGA module. As shown in Fig. 4, the CGA first sums the two inputs and then performs spatial attention and channel attention processing, respectively. These two pieces

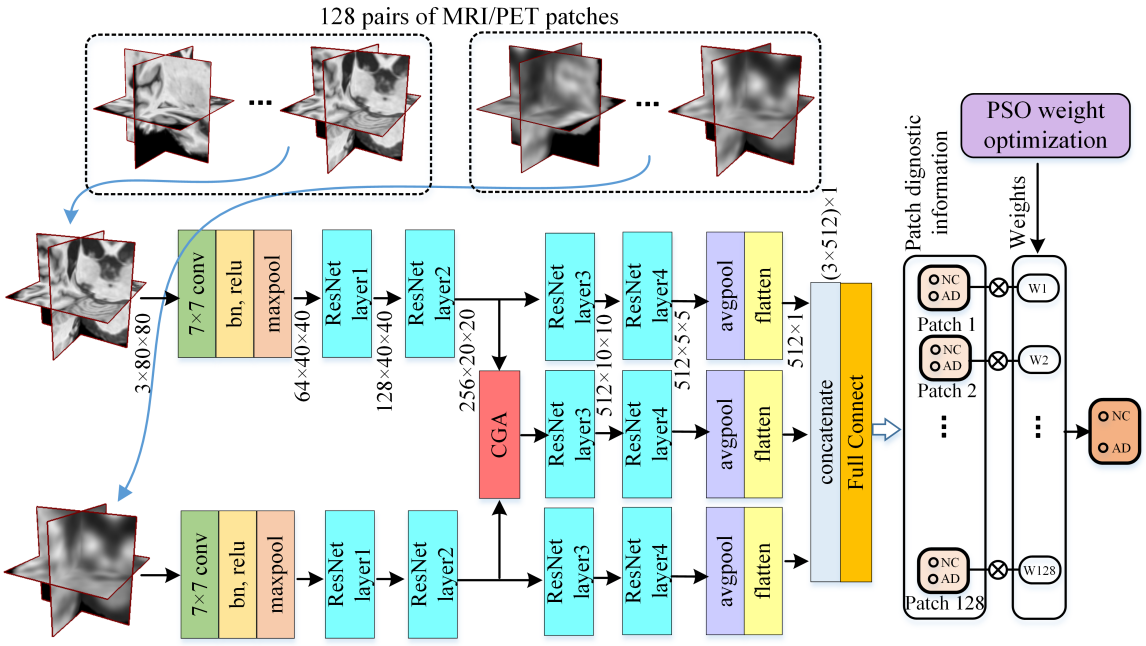

FIGURE 3. The architecture of multimodal 3-axial-plane patches fusion based framework for AD diagnose.

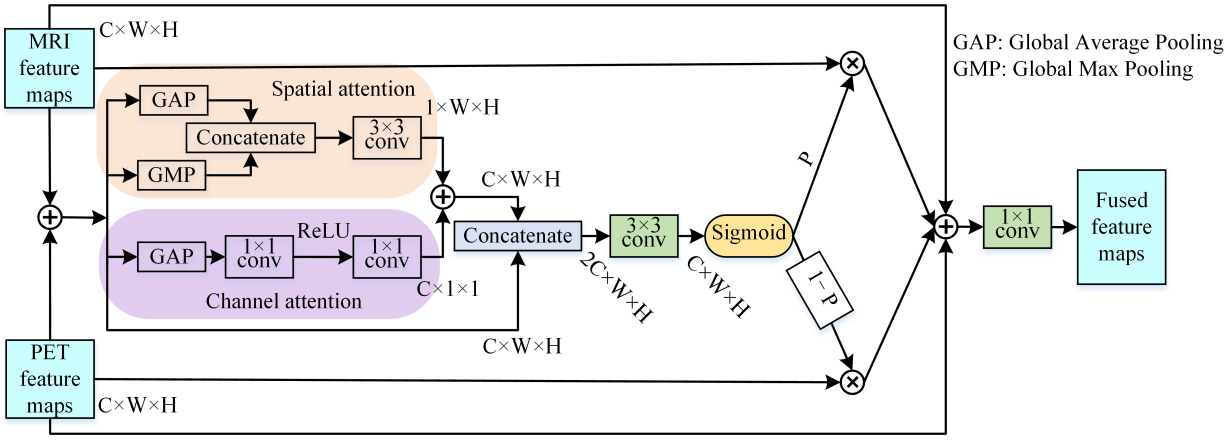

FIGURE 4. The architecture of multimodal 3-axial-plane patches fusion based framework for AD diagnose.

of attention information are channel-wise concatenate with features, and after processing using 3×3 convolution, the attention weight P is computed by Sigmoid function. The MRI features are multiplied by P, and the PET features are multiplied by $(1 - P)$. Finally, these two attention weighted features and the original features are merged as the output of the CGA.

The fused features output from the CGA module are processed by the two ResNet layers, then it is concatenated with the MRI and PET features as the third type feature. These multimodal features are then processed by the fully connected layer to be transformed into classification information as the output of the neural network.

2.3. Patch Weights Optimization with PSO Algorithm

After being processed by the multimodal fusion network, the output of each MRI/PET pair of patches can be used for the diagnosis of AD. And the diagnostic information from 128 patches can be combined to diagnose AD more precisely. However, the locations of these 128 patches are randomly selected from ROI, so the contributions of different patches may not be the same. Although combining them by same coefficients is a feasible method, a better way to combine them is by weighting coefficients. To find the optimal weight coefficients, PSO algorithm is adopted.

PSO is a stochastic search algorithm based on group collaboration [20], and the basic idea is to find the optimal solution through the collaboration and information sharing between individuals in the group. PSO firstly initializes a group of random particles (random solution) and then iteratively updates these particles to finally find the optimal solution. The steps of the PSO algorithm are as follows.

(1) Initialization: Initialize 100 particles, and each particle is a randomly generated 1×128 weight coefficient vector (particle position) with weights in the range of [0, 1]. In addition, each particle has a velocity attribute that represents the direction and magnitude of the position updating. Meanwhile, the AD diagnostic accuracy of the merged 128 patches with these weights is used as the optimization function to calculate the fitness value of these particles.

(2) Iterative search: In each iteration, the particle updates itself by tracking two best fitness values. The first is the optimal solution found by the particle itself, called the personal best (pBest). The other is the optimal solution found by the entire group, called the global best (gBest). The particle adjusts its velocity and position according to these two values, thus continuously approaching the optimal solution.

(3) Stop searching: When the iterations reaches a preset maximum value (which is 200 in this work), or minimum error criteria are attained, the PSO stops iterative search and outputs the best particle as the optimal solution.

The following pseudo-code in Algorithm 1 explains the PSO algorithm in more detail.

Algorithm 1 PSO algorithm for patches weights optimization.

Input: Patch classification information output of neuronal network
Output: Best particle (best weights) p_{gbest}
Settings: $particle\_number$ =200, $particle\_dimension$ = 128, $weight\_range$ = [0, 1], $maximum\_iterations$ = 100, $optimal\_function$ = AD diagnose accuracy;
for each particle p_i **do**
  for each dimension d **do**
    initialize position (or weight) w_{id} randomly within $weight\_range$
    initialize velocity v_{id} randomly
  end for
end for
iteration $k = 1$
while $k <= maximum\_iterations$ **do**
  for each particle p_i **do**
    fitness value $f_i = optimal\_function(p_i)$
    if $f_i > f_{i,pbest}$ **then**
      $f_{i,pbest} = f_i$
      $p_{i,pbest} = p_i$
    end if
  end for
  $p_{gbest} = p_{i,pbest}$ of the particle have the best fitness value $f_{i,pbest}$
  for each particle p_i **do**
    for each dimension d **do**
      update velocity: $v_{id} = v_{id} + c_1 r_1 (w_{id,pbest} - w_{id}) + c_2 r_2 (w_{d,gbest} - w_{id})$
      update position: $w_{id} = w_{id} + v_{id}$
    end for
  end for
  $k++$
end while

3. EXPERIMENTS AND RESULTS

3.1. Experimental Settings

The data used in this work includes 2031 samples, so there are 2031 pairs of MRI and PET images, which belong to 945 participants. These data were randomly divided into training set, validation set, and test set in the ratio of 70%, 10%, and 20%. To prevent data leakage, which may result from the samples of the same participants appearing in different data sets, the data set was divided based on participants, making sure that the samples from same participant would only appear in the same data set.

The patches from training set were firstly shuffled and then fed into neuronal network model during the training phase. The selection of hyperparameters was conducted through an empirical process involving multiple rounds of testing. Specifically, the key parameters such as learning rate, batch size, and dropout rates were adjusted based on their impact on model performance. Each configuration was evaluated using a validation set, and the final hyperparameters were chosen based on their greatest performance over multiple trials, ensuring robustness and minimizing the risk of suboptimal results. Finally, the model was trained in 120 epochs adopting Adam optimizer with the batch size of 256, and the learning rate is set to 0.0001. In the last 40 epochs, the performance is verified using the validation set for each epoch, and the model with the best performance was saved. When the training was over, the saved model would be used in the test set for performance testing.

The model was trained on an NVIDIA GeForce GTX 3090 GPU, and the THOP package was used to profile the proposed model. The number of parameters is 31.1 M; the size of model is 118.65 MB; and the computational complexity is 507.34 GFlops. During training phase, it took 134 seconds to train an epoch in GTX 3090 GPU. In the validation phase, it took only 0.035 seconds to process one sample.

The patches from the validation and test sets were processed through the trained model, and each sample would get diagnostic information of 128 patches. In order to optimize the weights of these 128 patches' information using PSO algorithm, the samples of validation set were used for the training of PSO algorithm. The PSO algorithm was set up with 100 particles, and the optimized weight coefficients were outputted after 200 iterations. These weight coefficients were used for merging the patch diagnostic information of the test set, and the diagnostic results of the test set would be attained. The performance of the model was verified by comparing the diagnostic results with the actual diagnostic results of test set. The experiment was conducted 10 times using repeated random-split validation, where the dataset was randomly divided into training, validation, and test sets in each iteration to ensure independence between iterations. The final performance metrics were calculated as the mean $\pm$ standard deviation across all 10 trials, providing a measure of model performance and stability.

3.2. Evaluation Criteria

To evaluate the performance of AD diagnostics, several commonly used metrics are adopted as evaluation criteria, including

TABLE 2. The AD diagnosis performances.

classification	ACC (%)	SEN/REC (%)	SPE (%)	PRE (%)	F1 (%)	AUC (%)
NC vs. AD	94.03 ± 0.97	91.85 ± 2.01	96.22 ± 0.9	96.06 ± 0.86	93.89 ± 1.04	97.22 ± 0.45
NC vs. MCI	71.02 ± 2.91	70.73 ± 13.62	71.33 ± 10.26	73.99 ± 3.88	71.22 ± 6.48	75.72 ± 2.05
MCI vs. AD	81.52 ± 3.75	84.37 ± 4.87	79.02 ± 10.31	79.03 ± 7.73	81.15 ± 2.53	88.43 ± 1.0

accuracy (ACC), sensitivity (SEN), specificity (SPE), precision (PRE), recall (REC, equivalent to SEN), F1-score (F1), and the area under the receiver operating characteristic curve (AUC). Assuming that *TP*, *FP*, *FN*, and *TN* denote the true positive, false positive, false negative, and true negative, the definitions of these metrics can be as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$SPE = \frac{TN}{FP + TN}, \quad SEN = \frac{TP}{TP + FN} \tag{2}$$

$$PRE = \frac{TP}{TP + FP}, \quad REC = \frac{TP}{TP + FN} \tag{3}$$

$$F1 = 2 \times \frac{PRE \times REC}{PRE + REC} \tag{4}$$

3.3. Results of AD Diagnosis

In order to demonstrate the contribution of each step in this methodology to the improvement of AD diagnostic accuracy, experiments by adding different modules step by step were conducted. As can be seen in Fig. 5, the accuracy of AD diagnosis improved as different steps were added to the framework. In Fig. 5, the first two results indicate that when only MRI and PET data were used, the AD diagnostic accuracy was 87.85% and 90.29%, respectively, which shows that the accuracy using PET images is higher than MRI images. The third result (labeled as "Concat") was gained when a two-way ResNet network was adopted for AD diagnosis, and the features of MRI and PET were concatenated before the fully connected layer. It can be seen that the diagnostic accuracy was improved to 91.88% because of the use of two modal data information. Subsequently, the CGA module was introduced into the neural network model for effective fusion of the two modalities through the attention mechanism, which have the performance as the fourth result (labeled as "+CGA") in Fig. 5, and it can be seen that the CGA further improved the diagnostic accuracy to 93.18%. Finally, with the addition of PSO algorithm, the weighting coefficient of multiple patches was optimized, and the diagnostic accuracy reached 94.03%, as shown in the last result (labeled as "+PSO") in Fig. 5. From these step-by-step experiments, it can be seen that each step used in this method plays an obvious role in the improvement of AD diagnostic performance.

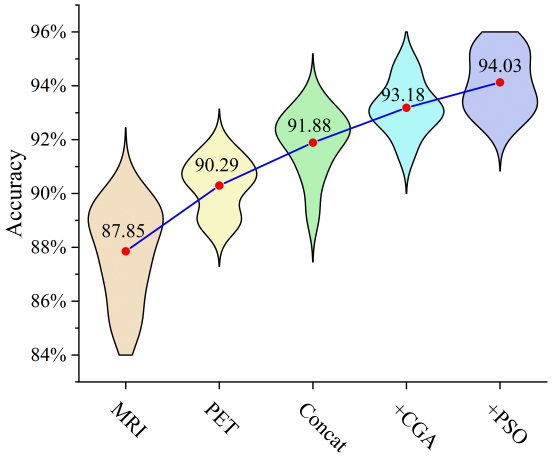

FIGURE 5. The contributions of each step.

The experiment results of various metrics of AD diagnosis are listed in Table 2, from which it can be seen that except the accuracy of 94.03%, the sensitivity (recall), specificity, and precision are 91.85%, 96.22%, and 96.06%, respectively, while the value of F1-score is 93.89%, and especially, the AUC is as high as 97.22%. These metrics are all indicators of the promising performance of the proposed method for AD diagnosis. In addition to the experiments of NC vs. AD, we also conducted experiments to categorize NC vs. MCI and MCI vs. AD. MCI is a prodromal stage of AD, and it is difficult to diagnose MCI by neuroimaging because the symptoms of this stage are not obvious, and the effect on the brain is very mild. As can be seen from the results of the experiment, the accuracy of these two classifications is also relatively low, amounting to 71.02% and 81.52%. While F1-scores were 71.22% and 81.15%, and AUCs were 75.72% and 88.43%, respectively. The receiver operating characteristic (ROC) curves for NC vs. AD, NC vs. MCI, and MCI vs. AD are also shown in Fig. 6.

3.4. Comparison with Other Studies

We have also evaluated the proposed approach against prior AD diagnosis studies. The comparison results are listed in Table 3. From these results, it can be found that the proposed method surpasses other methods in accuracy, F1 score, and AUC. There is at least a 2.1% improvement in accuracy compared with [5] and a 2.7% improvement in F1-score compared with [15], and compared to the method of [22], the present method also has 1% higher AUC. These results demonstrate the promising AD diagnosis performances of the proposed method compared to other methods.

TABLE 3. Comparison with other studies on AD diagnosis.

Studies	Methods	ACC (%)	F1 (%)	AUC (%)
[15]	Multi-modal cross-attention neuronal network	91.07	91.11	94.07
[5]	Artificial neural network with hybrid meta-ROI	91.9	-	-
[22]	Multi-modal feature selection algorithm with anchor graph	86.33	-	93.76
[21]	Spatially-constrained Fisher representation network	91.44	90.24	96.26
this study	Patch fusion neuronal network with PSO patch optimization	94.03	93.89	97.22

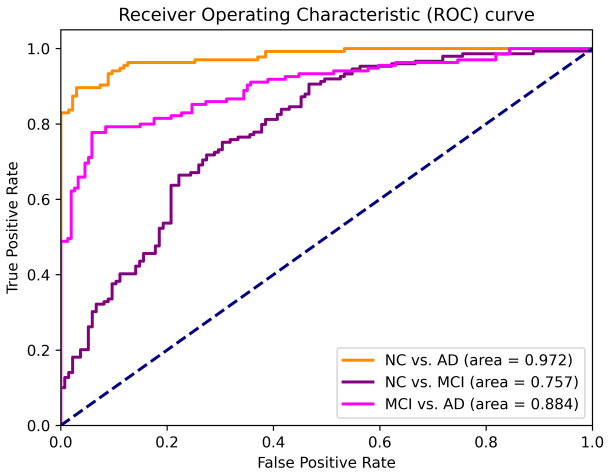

FIGURE 6. The receiver operating characteristic curves of AD diagnosis.

4. DISCUSSION

In this study, multiple 3-axial-plane patches were extracted from 3D MRI/PET images, and a multimodal fusion neural network was designed to process these patches. PSO algorithm was employed to optimize the patch weights to merge the diagnostic information of multiple patches, and promising diagnostic results were obtained. Usually, neuroimaging such as MRI/PET is 3D images, and processing 3D images takes a long time and a lot of computational power. When using neural network models to process 3D images, complicated neural network architectures and advanced algorithms usually cannot be applied due to the limited computational power and graphic memory. Moreover, the number of samples of 3D images is usually not large enough to adequately train the neural network model, resulting in insufficient generalization ability of the model. In this study, benefited from the patch extraction method, a deeper and complex neural network model and attention approach can be used for multimodal data fusion, and multiple patches can be extracted for each 3D image sample, which serves as data augmentation, and therefore effectively enhances the efficiency of multimodal fusion and improves the accuracy of AD diagnosis.

From the experiment results of AD diagnosis, it can be found that a high accuracy can be achieved using neuroimaging for the classification of NC vs. AD; however, the accuracy for the classification of NC vs. MCI and MCI vs. AD is relatively low, especially for NC vs. MCI, which has the lowest accuracy rate. We believe that this is because MCI is the prodromal stage of AD, and the lesions in the patient's brain caused by the disease are relatively mild. These mild lesions are difficult to differentiate by neuroimaging, and therefore the diagnosis of NC and MCI by neuroimaging is less accurate. As the disease progresses, the lesions become more severe, and the difference in neuroimaging becomes more obvious, so the classification of MCI vs. AD is somewhat more accurate than that of NC vs. MCI.

In the design of the AD diagnostic method in this work, we focus on the use of neuroimaging in the diagnosis of AD, so only two types of image data, MRI and PET, are utilized. There are actually other non-imaging data that can be used for AD diagnosis, such as aberrant levels of Beta-amyloid and Tau in cerebrospinal fluid are biomarkers of AD, making Beta-amyloid and Tau buildup useful information for the diagnosis of AD [23]. Other AD-related risk factors including APOE4 gene and age are also helpful in the diagnosis of AD [24]. If these non-imaging data are added into the diagnostic approach, it will certainly further improve the diagnosis of AD. However, just concatenating these data is not enough, if the neural network can be carefully designed, and methods such as cross-attention are used to improve the fusion effect of image and non-image data, the improvement of the diagnostic accuracy of AD will be even greater, and this is also our future research direction.

In the proposed framework, for the optimization of the weight coefficients for the merging of patches, we use PSO algorithm. Of course, the weight coefficients can also be optimized using other optimization algorithms such as Adam and SGD with momentum. However, we found that PSO algorithm, although not a gradient-based optimization, can achieve the same performance as Adam and slightly better than SGD. This observation may stem from task-specific characteristics: in parameter spaces with moderate dimensionality and parallelizable computational resources, PSO's swarm collaboration effectively balances exploration and exploitation, mitigating over-reliance on singular gradient directions. Furthermore, PSO's elimination of manual learning rate tuning reduces hyperparameter calibration costs. Thus, for optimization scenarios for our task, PSO proves to be an efficient and feasible choice. Future work may investigate dynamic optimizer switching strategies to address more complex task requirements.

In this study, MMSE and CDR scores were not included as features because they are highly association with the labels. These metrics serve as established diagnostic criteria for AD in clinical practice, so the diagnose results in ADNI, which are used as labels in the training and validation of model, mainly come from MMSE and CDR. Although integrating these scores as predictive features could improve model performance, such practice risks introducing diagnostic bias and overestimated findings for the model. Therefore, in this work, we did not include MMSE and CDR scores as feature sets.

5. CONCLUSION

In this work, a neural network model for multimodal patch fusion based on the attention mechanism is proposed. Multiple 3-axial-plane patches are extracted from the coronal, transverse, and sagittal planes of 3D MRI and PET images, and then fed into a neural network model, which is based on a two-way Resnet-18 and content-guided attention module, for multimodal feature fusion. After processing by the neural network model, each patch gets AD diagnostic information. In order to merge these pieces of diagnostic information of patches efficiently, PSO algorithm is used to find the optimal patch weight coefficients. Eventually, the experiment results show that the merged information can diagnose AD with promising accuracy of 94.03% and demonstrates the effectiveness of the proposed method.

ACKNOWLEDGEMENT

This work was supported by Fujian Provincial Natural Science Foundation of China [grant number 2022J011271, 2023I0044].

REFERENCES

[1] Mank, A., J. J. M. Rijnhart, I. S. Van Maurik, L. Jönsson, R. Handels, E. D. Bakker, C. E. Teunissen, B. N. M. Van Berckel, A. C. Van Harten, J. Berkhof, and W. M. Van Der Flier, "A longitudinal study on quality of life along the spectrum of Alzheimer's disease," *Alzheimer's Research & Therapy*, Vol. 14, No. 1, 132, Sep. 2022.

[2] Rao, Y. L., B. Ganaraja, B. V. Murlimanju, T. Joy, A. Krishnamurthy, and A. Agrawal, "Hippocampus and its involvement in Alzheimer's disease: A review," *3 Biotech*, Vol. 12, No. 2, 55, Feb. 2022.

[3] Chandra, A., G. Dervenoulas, and M. Politis, "Magnetic resonance imaging in Alzheimer's disease and mild cognitive impairment," *Journal of Neurology*, Vol. 266, 1293–1302, Jun. 2019.

[4] Moradi, E., A. Pepe, C. Gaser, H. Huttunen, and J. Tohka, "Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects," *Neuroimage*, Vol. 104, 398–412, Jan. 2015.

[5] Zheng, X., "Detection of Alzheimer's disease using hybrid meta-ROI of MRI structural images," *Diagnostics*, Vol. 14, No. 19, 2203, Oct. 2024.

[6] Al Shehri, W., "Alzheimer's disease diagnosis and classification using deep learning techniques," *PeerJ Computer Science*, Vol. 8, e1177, Dec. 2022.

[7] Ghosh, T., M. I. A. Palash, M. A. Yousuf, M. A. Hamid, M. M. Monowar, and M. O. Alassafi, "A robust distributed deep learning approach to detect Alzheimer's Disease from MRI images," *Mathematics*, Vol. 11, No. 12, 2633, Jun. 2023.

[8] Liu, Y., S. Mazumdar, and P. A. Bath, "An unsupervised learning approach to diagnosing Alzheimer's disease using brain magnetic resonance imaging scans," *International Journal of Medical Informatics*, Vol. 173, 105027, May 2023.

[9] Chételat, G., J. Arbizu, H. Barthel, V. Garibotto, I. Law, S. Morbelli, E. Van De Giessen, F. Agosta, F. Barkhof, D. J. Brooks, *et al.*, "Amyloid-PET and 18F-FDG-PET in the diagnostic investigation of Alzheimer's disease and other dementias," *The Lancet Neurology*, Vol. 19, No. 11, 951–962, Nov. 2020.

[10] De Santi, L. A., E. Pasini, M. F. Santarelli, D. Genovesi, and V. Positano, "An explainable convolutional neural network for the early diagnosis of Alzheimer's disease from 18F-FDG PET," *Journal of Digital Imaging*, Vol. 36, No. 1, 189–203, Nov. 2023.

[11] Bi, S., S. Yan, Z. Chen, B. Cui, Y. Shan, H. Yang, Z. Qi, Z. Zhao, Y. Han, and J. Lu, "Comparison of 18F-FDG PET and arterial spin labeling MRI in evaluating Alzheimer's disease and amnestic mild cognitive impairment using integrated PET/MR," *EJNMMI Research*, Vol. 14, No. 1, 9, Jan. 2024.

[12] Frings, L., G. Blazhenets, J. Brumberg, A. Rau, H. Urbach, and P. T. Meyer, "Deformation-based morphometry applied to FDG PET data reveals hippocampal atrophy in Alzheimer's disease," *Scientific Reports*, Vol. 14, No. 1, 20030, Aug. 2024.

[13] Beheshti, I., N. Geddert, J. Perron, V. Gupta, B. C. Albensi, and J. H. Ko, "Monitoring alzheimer's disease progression in mild cognitive impairment stage using machine learning-based FDG-PET classification methods," *Journal of Alzheimer's Disease*, Vol. 89, No. 4, 1493–1502, 2022.

[14] Hojjati, S. H. and A. Babajani-Feremi, "Prediction and modeling of neuropsychological scores in Alzheimer's disease using multimodal neuroimaging data and artificial neural networks," *Frontiers in Computational Neuroscience*, Vol. 15, 769982, Jan. 2022.

[15] Zhang, J., X. He, Y. Liu, Q. Cai, H. Chen, and L. Qing, "Multi-modal cross-attention network for Alzheimer's disease diagnosis with multi-modality data," *Computers in Biology and Medicine*, Vol. 162, 107050, Aug. 2023.

[16] Salahuddin, Z., H. C. Woodruff, A. Chatterjee, and P. Lambin, "Transparency of deep neural networks for medical image analysis: A review of interpretability methods," *Computers in Biology and Medicine*, Vol. 140, 105111, Jan. 2022.

[17] Litjens, G., T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, Vol. 42, 60–88, Dec. 2017.

[18] Xu, W., Y.-L. Fu, and D. Zhu, "ResNet and its application to medical image processing: Research progress and challenges," *Computer Methods and Programs in Biomedicine*, Vol. 240, 107660, Oct. 2023.

[19] Chen, Z., Z. He, and Z.-M. Lu, "DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention," *IEEE Transactions on Image Processing*, Vol. 33, 1002–1015, 2024.

[20] Piotrowski, A. P., J. J. Napiorkowski, and A. E. Piotrowska, "Particle swarm optimization or differential evolution — A comparison," *Engineering Applications of Artificial Intelligence*, Vol. 121, 106008, May 2023.

[21] Pan, Y., M. Liu, C. Lian, Y. Xia, and D. Shen, "Spatially-constrained fisher representation for brain disease identification with incomplete multi-modal neuroimages," *IEEE Transactions on Medical Imaging*, Vol. 39, No. 9, 2965–2975, Sep. 2020.

[22] Li, J., H. Xu, H. Yu, Z. Jiang, and L. Zhu, "Multi-modal feature selection with anchor graph for Alzheimer's disease," *Frontiers in Neuroscience*, Vol. 16, 1036244, Nov. 2022.

[23] Papaliagkas, V., K. Kalinderi, P. Vareltzis, D. Moraitou, T. Papamitsou, and M. Chatzidimitriou, "CSF biomarkers in the early diagnosis of mild cognitive impairment and Alzheimer's disease," *International Journal of Molecular Sciences*, Vol. 24, No. 10, 8976, May 2023.

[24] Sun, Y.-Y., Z. Wang, and H.-C. Huang, "Roles of apoE4 on the pathogenesis in Alzheimer's disease and the potential therapeutic approaches," *Cellular and Molecular Neurobiology*, Vol. 43, No. 7, 3115–3136, May 2023.