



A Multitask Neural Network Framework for Multistage Alzheimer's Disease Diagnosis and Hippocampal Segmentation Based on Neuroimaging Modalities

Fuchun Jiang^{1,2*}, Xinjie Yang², Weiming Lin^{1,2}, Rongzhu Zou³, Bing Luo³

¹ School of Opto-Electronic and Communication Engineering, Xiamen University of Technology, Xiamen 361024, China

² Fujian Key Laboratory of Communication Network and Information Processing, Xiamen University of Technology, Xiamen 361024, China

³ Xiamen Municipal Utility Tunnel Investment & Management Co., Ltd, Xiamen 361000, China

Corresponding Author Email: jiangfuchun@xnut.edu.cn

Copyright: ©2025 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.420635>

ABSTRACT

Received: 19 May 2025

Revised: 13 October 2025

Accepted: 29 October 2025

Available online: 31 December 2025

Keywords:

AD, MCI, neuroimaging, deep convolutional neural network, multitask learning, hippocampal segmentation, multiclass diagnosis, ELM, feature integration, PCA

Alzheimer's disease (AD) is a progressive neurodegenerative disorder, with mild cognitive impairment (MCI) representing an intermediate stage between normal cognitive function (NC) and AD. Accurate multistage diagnosis of AD remains a significant challenge due to the lack of distinct boundaries between adjacent stages. This study introduces classification and segmentation multitask network (CSMT-Net), a novel deep learning framework designed to address these challenges by simultaneously performing classification and segmentation tasks for AD diagnosis. The proposed network extracts features from neuroimaging modalities, including magnetic resonance imaging (MRI) and positron emission tomography (PET), and employs a multitask approach to learn AD-related pathology through both classification and structural segmentation of the hippocampus, a critical biomarker of AD. Principal component analysis (PCA) is applied to the extracted features for dimensionality reduction and feature selection, which is further integrated with other diagnostic information, such as cerebrospinal fluid (CSF) biomarkers, genetic factors, age, gender, and educational background. This integrated feature set is used for multiclass diagnosis of AD. An extreme learning machine (ELM) is employed as the classifier to predict the likelihood of AD across multiple stages. Evaluation on the AD Neuroimaging Initiative (ADNI) dataset shows that the CSMT-Net framework achieves an accuracy of 69.3% and an F1-score of 69.7% in the multistage diagnosis of AD. The results indicate that the multitask approach significantly enhances diagnostic accuracy compared to single-task methods. The integration of both classification and segmentation tasks within the CSMT-Net framework demonstrates its potential to improve the precision of multistage AD diagnosis, offering a promising tool for advancing clinical diagnostic capabilities.

1. INTRODUCTION

AD is one of the causes of dementia, which is a brain disorder caused by damage to nerve cells in the brain, leading to mental decline and memory loss. AD is a progressive disease, as the nerve cells are damaged, the patient gradually experiences changes in memory, language, mood, personality or behavior, eventually affecting basic bodily functions such as walking and swallowing, and can be life-threatening [1, 2]. Studies have shown that the average survival time after diagnosis is 4-8 years [3]. The onset of AD can be divided into three stages: NC, MCI due to AD, and dementia due to AD. Patients with MCI often exhibit mild brain lesions, but do not show significant symptoms. Approximately one-third of MCI patients will progress to AD within five years. There is no effective way or drug to treat AD, but some studies have shown that some patients with MCI do not have additional cognitive decline or revert to normal cognition [4]. Preventive treatment at the MCI stage can be effective in slowing or stopping the AD progress. Thus, a multistage diagnosis of AD can assist in differentiating these three groups, which is helpful

in determining appropriate therapeutic interventions.

Neuroimaging data can be employed as diagnostic data in computer-aided diagnostic techniques for AD. For instance, AD-related neuronal loss results in anatomical alterations in the patient's brain tissue that can be identified by structural MRI scans [5, 6]. For this reason, MRI is a frequently used image data set. Fluorodeoxyglucose positron emission tomography (FDG-PET) is another medical imaging modality that can identify abnormalities in glucose metabolism caused by AD. As a result, the distinctive pattern of electron glucose metabolism in PET images can be used to distinguish between patients and NC [7, 8]. Furthermore, aberrant levels of beta-amyloid and tau in CSF are established biomarkers of AD, making beta-amyloid and tau buildup useful information for the diagnosis of AD [9]. While these medical data can independently diagnose AD, integrating multimodal data significantly enhances diagnostic precision [10, 11].

Currently, the binary diagnosis of AD has achieved a high accuracy rate, however, the multistage diagnosis task, which can distinguish AD, MCI and NC simultaneously, has more practical value in clinical applications. But the accuracy of

computer-aided multistage diagnosis of AD is still low [12-14] due to the slight variations in the pathogenesis of AD between adjacent stages, particularly in the early stages of MCI where the pathological changes are very small in comparison to normal individuals. In order to more precisely distinguish the different stages of AD, multitask deep learning models have been used to improve diagnosis accuracy [15, 16], i.e., to acquire the diagnostic results while also outputting additional metrics, such as mini-mental state examination (MMSE) and clinical dementia rating (CDR). Given that AD results in hippocampal atrophy, and the hippocampal regions are considered to be the main landmark areas for lesions in Alzheimer's studies [17, 18], allowing the neural network model to learn the structural features of the hippocampus will improve the ability to diagnose AD of different stage. This paper proposes a novel multitask deep neural network model named Classification and Segmentation Multitask Network (CSMT-Net), which adds a segmentation task to segment the hippocampal structure along with the MMSE, CDR regression and AD classification tasks. The experiments of multistage diagnosis (AD vs. MCI vs. CN) were conducted. Based on the results of the experiments, the accuracy of AD multistage diagnosis can be further enhanced by including the hippocampus segmentation task.

The main highlights of this paper are as follows:

- 1) This work proposed a deep learning-based approach for AD multistage diagnosis. A multitask deep neural network named CSMT-Net was designed for AD-related feature extraction with neuroimaging, such as MRI and FDG-PET.
- 2) Hippocampus segmentation was utilized as one task of CSMT-Net for structural MRI, so that the network could learn about the AD-related structural lesion information, which could indicate the progress of AD and is useful for AD multistage diagnosis.
- 3) A machine learning framework consisting of PCA and ELM was developed for AD multistage diagnosis using multimodal data, including the neuroimaging features produced by CSMT-Net and additional clinical data, such as CSF biomarkers, genetic factors, and demographic information.

2. RELATED WORKS

Using artificial intelligence algorithms based on neuroimaging and clinical data, computer-aided diagnosis of AD involves grouping individuals into categories. Classification approaches include binary classification [19, 20], multistage classification [13, 21], and AD onset prediction [22, 23]. Binary classification involves distinguishing NC vs. AD, NC vs. MCI, and MCI vs. AD. On the other hand, multistage classification can categorize persons as NC, MCI, or AD simultaneously, which is more suitable for clinic applications. When predicting the onset of AD, MCI is categorized as progressive MCI, which is likely to develop AD within a few years, and stable MCI, which is not expected to proceed to AD. This prediction process essentially involves a binary classification task. Among these AD diagnostic approaches, there are mainly machine learning and deep learning techniques.

2.1 Machine learning-based AD diagnosis approach

Machine learning techniques provide a methodical way to

build sophisticated, automatic classification models to handle massive volumes of data and identify subtle and complicated patterns. Establishing the architectural design is necessary when applying these machine learning techniques to the classification of AD. Four stages are usually needed: feature extraction, feature selection, dimension reduction, and classification algorithm implementation. Numerous machine learning algorithms have demonstrated efficacy in the classification of AD. For example, Dong et al. [24] proposed a latent feature fusion-based technique to utilize the information contained in multimodal image data. They developed a unique projection matrix for every modality, after which they projected and fused latent feature representations of several modalities onto a low-dimensional target space for AD classification. Feng et al. [25] suggested using an ROI-based contourlet sub-band energy feature to represent the MRI image in the frequency domain. Sub-band energy feature vectors were created from 90 ROIs in order to record their contour data and energy distribution, then these features were concatenated and fed into support vector machine (SVM) for AD classification. Zhou et al. [26] developed a machine learning-based segmentation and classification pipeline for AD classification. They firstly segment the hippocampus from MRI. Then selected 37 features most relevant to AD by the hierarchical clustering method and least absolute shrinkage and selection operator algorithm. Ultimately, four classifiers were used with selected features to differentiate AD from NCs. Although machine learning-based models can diagnose AD efficiently, most of them are unable to extract adaptive characteristics, so they typically need human-generated features.

2.2 Deep learning-based AD diagnosis approach

Deep learning has grown quickly in recent years because of the increasing GPU processing capacity, and because it does away with the necessity for manual feature extraction, it is now frequently employed in medical image-aided diagnosis applications. Thus, it became feasible to classify different AD stages using deep learning models [27]. For example, an ensemble model based on a 3-D convolutional neural network and genetic algorithm is proposed in study [28], which can differentiate the subjects with AD or MCI and also identify the discriminative brain regions significantly contributing to the classifications. In study [29], a novel two-stage deep learning AD progression detection framework was proposed, this method utilized information fusion of different patient longitudinal multivariate modalities, so it can predict the precise AD onset time of MCI patients. A multiclass classification task was utilized in the first stage to estimate a patient's diagnosis, and a regression task was used in the second stage to predict the precise conversion time of patients with MCI. Wang et al. [12] presented an asymmetry-enhanced attention network for AD diagnosis, which proficiently integrates the cerebral anatomical asymmetry properties to enhance the accuracy and stability of classification tasks.

2.3 Multitask deep neural network for AD diagnosis

In order to increase the learning efficiency of the models, some deep learning approaches use a multitask strategy to train the neural network, where the tasks are typically related to AD. This can force the neural network model to learn more AD-related information from the data, improving the accuracy of the classification task. For example, in study [30], a deep

multitask multi-channel learning framework was developed for simultaneous brain illness categorization and clinical score regression, using MRI data and demographic information. Dong et al. [31] used a pre-trained deep model as a feature extractor to generate high-level feature maps of different tasks. However, segmentation has not been studied as a task in the multitask AD diagnostic neural network models. Since the degree of AD condition can be indicated by lesions in brain neural structures on MRI, improving the accuracy of AD classification can be facilitated if a neural network can simultaneously learn to recognize the structural knowledge of regions related to AD lesions in MRI images. Therefore, to improve the AD diagnosis accuracy, this paper examined the possibility of using multitask neural networks to include hippocampal segmentation as one of the tasks. Then, the trained deep neural networks will be utilized to extract AD-related features from multimodal neuroimaging, and these multimodal features will then be fed into a machine learning model. This approach will effectively use the benefits of multimodality and multitasking to improve the accuracy of

AD classification.

3. METHODS

In this section, a multimodal framework for AD multistage diagnosis is introduced, Figure 1 displays the architecture of this framework. The framework comprises two components: the CSMT-Net part and a machine learning part. The CSMT-Net, which is pre-trained with classification, regression and segmentation tasks, extracts deep features from preprocessed 3D MRI and 3D PET images. Subsequently, the deep features of MRI and PET are reduced to 5 features each using the PCA algorithm. The neuroimaging deep features are then concatenated with hippocampus volume, CSF biomarkers, apolipoprotein E4 (APOE4), a genetic risk factor for AD [32], and demographic data including age, gender, and education, before an ELM classifier is employed with these features to generate a multistage diagnosis.

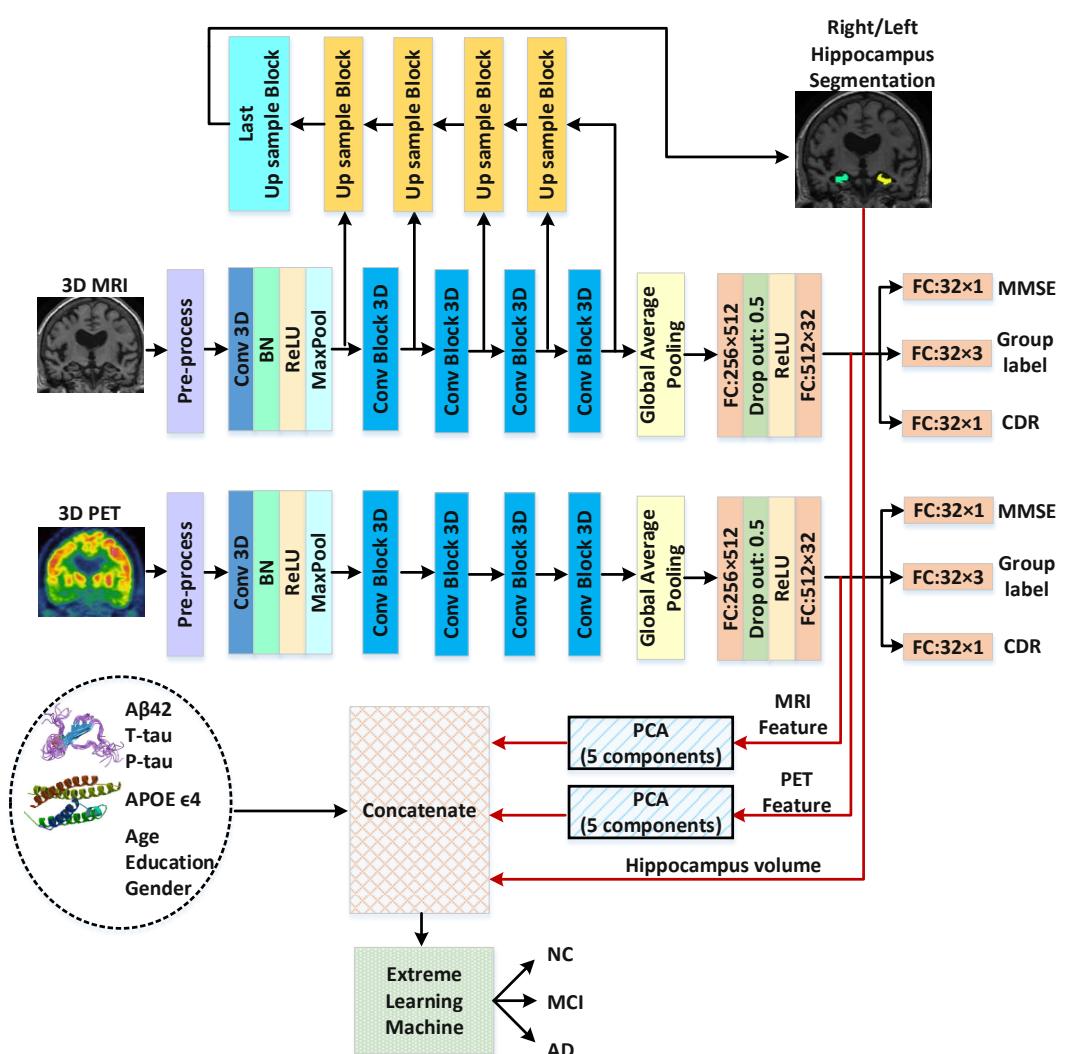


Figure 1. The overall framework of the CSMT-Net based multimodal AD multistage diagnosis

3.1 Neuroimaging preprocessing

The images in this paper were acquired from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<https://ida.loni.usc.edu/>). Due to variations in size, head position, and orientation of the MRI and PET images,

preprocessing of these neuroimages was required. Preprocessing involves skull stripping and rigid registration. The MRI images underwent N4 bias field correction with the ANTs tool, skull stripping with FSL software, rigid registration to the MNI152 template using the IRTK tool, and cropping to produce a $128 \times 160 \times 128$ -sized 3D brain image.

The PET images were initially registered with the raw MRI image, skull stripped using the MRI skull stripping mask, aligned with the registered MRI, and then cropped to a size of $128 \times 160 \times 128$. Data augmentation techniques were utilized during the neural network training phase to expand the training dataset. This involved randomly adjusting the cropping center point, flipping images along the left/right axis, and applying random rotations within a range of ± 20 degrees in three dimensions.

3.2 The architecture of the CSMT-Net

The architecture of CSMT-Net presented in this paper is illustrated in Figure 2. It begins with an input layer that processes the input image. This layer includes a 3D convolutional layer with a convolutional kernel size of $3 \times 3 \times 3$, followed by a batch normalization (BN) layer and a Rectified Linear Unit (ReLU) activation layer. Subsequently, maximum pooling is applied, which decreases the feature map size to $64 \times 80 \times 64$ and generates 16 feature maps. Following the input

layer, there are four 3D convolution blocks that utilize a residual structure. The 3D convolution block consists of three sets of convolutions, BN layer, and ReLU layer. The input is combined with the output of these convolutions through residual linking. The combined feature maps are then sent through another set of convolutions, BN layer, ReLU layer, and a maximum pooling layer. The feature map size is reduced by half in each 3D convolution block, while the number of feature maps is doubled. As a result, the fourth residual convolution block produces 256 feature maps with dimensions of $4 \times 5 \times 4$. A global average pooling operation is performed to calculate the average values of all feature maps, resulting in a 1×256 vector. A 256×512 fully connected layer is followed by a dropout layer, a ReLU activation layer, and another 512×32 fully connected layer to produce a 1×32 vector, which represents the deep features of the neuroimage. During the network training phase, this feature vector is utilized to create training targets for the multitask learning, including MMSE, CDR, and classification label, through two 32×1 fully connected layers and one 32×3 fully connected layer.

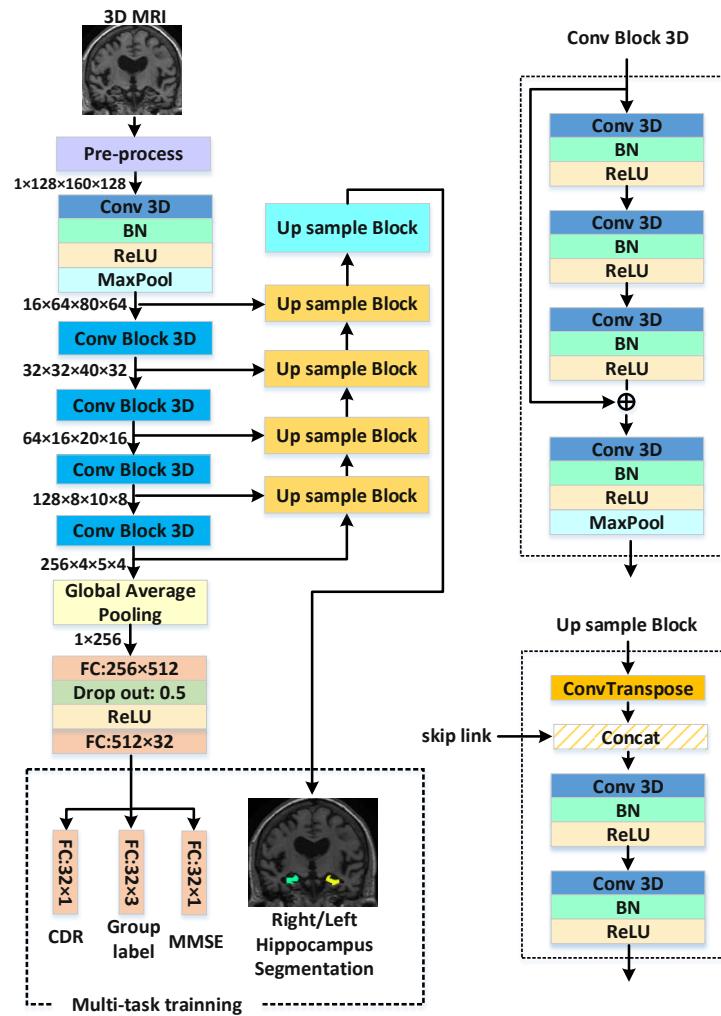


Figure 2. The architecture of the CSMT-Net

A U-Net-similar up-sampling branch is added to the backbone network for the purpose of segmenting the right and left hippocampus, for the network designed for MRI feature extraction, as depicted in Figure 2. The up-sampling module takes the input, up-samples it using an inverse convolution layer, and combines it with the relevant feature maps from the backbone network using skip links. The resulting output is

passed through two sets of convolutions, a BN layer, and a ReLU activation layer. The up-sampling module produces feature maps that are twice the size of the input. After four up-sampling modules, 16 feature maps of size $64 \times 80 \times 64$ are generated. These feature maps are further up-sampled by the final inverse convolution, resulting in segmentation outputs of size $128 \times 160 \times 128$. The hippocampus segmentation results

will be utilized as one of the tasks in the training of the MRI network. As manual segmentation of hippocampal labels is difficult to achieve, the segmentation tool in FSL software was utilized to segment the hippocampus from the MRI, and the segmentation results are then used as the training labels for the segmentation task. While training the MRI network for segmentation, the network can acquire segmentation skills and morphological knowledge similar to the FSL software, enhancing its capacity to extract valuable information from MRI data. The hippocampal segmentation output is also utilized to calculate the hippocampal volume, which can serve as an AD biomarker.

3.3 Loss function

When training the CSMT-Net, the loss function of the multitask training objective must be combined. The PET network has three tasks: multi-classification, regression of MMSE and CDR. A fourth task, which is segmenting the hippocampus, is included in the MRI network. The loss functions for these tasks are built as follows.

The multi-classification task utilized the multi-margin loss, which can enlarge distances of interclass and reduce intra-class variations simultaneously [33]. The loss function of multi-margin loss can be expressed as:

$$L_c = \frac{1}{C} \sum_{i=0 \& i \neq y_n}^{C-1} \max \left(0, w[y] * (\text{margin})^p \right) \quad (1)$$

where, C denotes the group number, specifically set to 3. $x[y]$ represents the output of the correct group, and $x[i]$ are the outputs of other groups. For the situation of sample imbalance, $w[y]$ is the weight for each group, which is set to 1. p and margin are set to the default value, which is 1.

For the tasks of MMSE and CDR regression, the loss function is squared loss, which can be expressed as:

$$L_{mmse} = (y_{mmse} - \hat{y}_{mmse})^2, \quad L_{cdr} = (y_{cdr} - \hat{y}_{cdr})^2 \quad (2)$$

where, y represents the output value, and \hat{y} represents the true value.

For the task of hippocampus segmentation, the binary cross-entropy loss for each voxel was calculated and averaged as the segmentation loss function, which can be expressed as:

$$L_{seg} = -\frac{1}{N} \sum_{i=1}^N y_i \log(p(y_i)) + (1-y_i) \log(1-p(y_i)) \quad (3)$$

where, N represents the number of voxels, y represents the ground truth label, and $p(y_i)$ represents the output probability of label.

Finally, we combine the above loss functions by weighting to get the final loss function as follows:

$$Loss = \frac{1}{4} \left(w_c \cdot L_c + w_{mmse} \cdot L_{mmse} + w_{cdr} \cdot L_{cdr} + w_{seg} \cdot L_{seg} \right) \quad (4)$$

where, $w_c, w_{mmse}, w_{cdr}, w_{seg}$ are the weighting coefficients of the four loss functions. Take note that the different loss functions often operate on vastly different scales, the individual loss

functions are scaled by these weighting coefficients to bring them into a comparable magnitude. This ensures that each task's contribution to the overall multitask loss is more balanced, preventing one task from overshadowing others purely due to its inherent scale. According to the output scale of each loss function, w_c was set to 1, w_{mmse} was set to 0.03, w_{cdr} was set to 0.1, w_{seg} was set to 1.

3.4 Classifier

The proposed method involves extracting MRI and PET features using a well-trained CSMT-Net that outputs 32 deep features. The features are initially processed using PCA to merge the highly correlated features [34]. Five major components are kept, while the rest are removed as noise. So 5 deep features are extracted from each MRI and PET image. These neuroimaging features are combined with additional biomarkers and demographic data, then inputted into a classifier for AD diagnosis. A classifier using ELM with Gaussian kernels [35] is created to do multistage classification of AD using multimodal data. The ELM algorithm can be described as follows: assuming there are N training samples $[x_1, x_2, \dots, x_N]$, in which, x_n represents the n -th sample consisting of M features. $Y \in \mathbb{R}^{N \times G}$ is a one hot ground truth label matrix for N samples of G classes. Upon receiving a new sample x , the label of x can be predicted as

$$f(x) = \begin{bmatrix} K(x, x_1) \\ K(x, x_2) \\ \vdots \\ K(x, x_N) \end{bmatrix} (\Omega + I/C)^{-1} Y \quad (5)$$

where, the variable C is a regularization coefficient set to 1. The variable γ is a parameter of the Gaussian kernel, which was set to 10 times of M in this study, and $K(x, x_n)$ is the Gaussian kernel described as:

$$K(u, v) = \exp(-\|u - v\|^2 / \gamma) \quad (6)$$

and Ω is an $N \times N$ kernel matrix that is calculated with N training samples:

$$\Omega = \begin{bmatrix} K(x_1, x_1) & K(x_1, x_N) \\ K(x_2, x_1) & \dots & K(x_2, x_N) \\ \vdots & & \vdots \\ K(x_N, x_1) & K(x_N, x_N) \end{bmatrix} \quad (7)$$

4. EXPERIMENTS AND RESULTS

4.1 Experiment data and implementation details

The data for this study were obtained from the ADNI dataset, which has recruited over 1800 participants aged 55 and above. Only participants with all necessary modalities were selected for validation, resulting in 263 NC, 299 MCI, and 263 AD samples. These 825 samples were collected from 560 participants, indicating that some samples were obtained from follow-up examinations of the same participant. It is necessary to note that there are samples of 53 participants who had changed groups during the follow-up study, that means these samples may come from the same participant but belong to

different groups, which made the classification more challenging. Aside from these 560 patients, the MRI scans of the remaining participants were used to train the CSMT-Net for MRI. The training dataset consisted of 6189 MRI scans, comprising 2080 NC, 2936 MCI, and 1173 AD. Similarly, 1600 PET images were utilized to train the network for PET, comprising 447 NC, 892 MCI, and 261 AD. During the network training phase, these 3D images underwent random cropping, flipping, and rotation in three axes by ± 20 degrees for data augmentation. The networks underwent training for 220 epochs. The initial 120 epochs were trained with a learning rate of 0.001, while the subsequent 100 epochs were trained with a learning rate of 0.0001. The complexity of the proposed neural network model is characterized by 4,449,344 parameters and a computational cost of 41.44 GFLOPs. For the training phase, an NVIDIA RTX 3090 GPU, equipped with 24GB of video memory, was utilized. Adopting a batch size of 4, the duration for each training epoch was approximately 850 seconds. Consequently, the entire training process, spanning 220 epochs, amounted to a total of about 52 hours.

Following the training of the CSMT-Net, deep features from neuroimaging data were extracted from the MRI and PET of the 825 samples previously described. During the validation phase, the MRI and PET features were initially processed using the PCA technique, and then all modality features were combined for classification validation. To prevent data leakage, the 825 samples collected from 560 participants were divided into training and testing datasets using 5-fold cross-validation, ensuring that samples from the same participants were assigned to either the training or testing dataset. Table 1 displays the demographic information of these 825 samples.

Table 1. The demographic information of the validation samples

	Number	Age	Gender (M/F)	Education (Years)	MMSE	CDR
NC	263	74.8 \pm 6.2	129/134	16.6 \pm 2.8	29.0 \pm 1.20.04 \pm 0.45	
MCI	299	73.8 \pm 7.5	155/144	16.2 \pm 2.8	27.5 \pm 2.0	1.48 \pm 1.0
AD	263	75.4 \pm 7.2	157/106	15.6 \pm 2.7	22.6 \pm 3.24.99 \pm 2.19	

We assessed the performance of multiclass classification by using accuracy, which measures the proportion of correctly categorized samples to the total samples, and F1-score. The F1-score is determined by combining precision (the ratio of correctly classified positive samples for one class to the total classified positive samples for one class) and recall (the ratio of correctly classified positive samples for one class to the real samples for one class) as follows:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

4.2 The results of AD classification

The CSMT-Net underwent training for 220 epochs, and the loss curves can be seen in Figure 3. Following network training and feature processing, 5-fold cross-validation was performed through a total of 100 runs, and the mean and standard deviation of these 100 runs were computed as the results. Experiments of binary classifications for NC vs. AD, NC vs. MCI, and MCI vs. AD, in addition to the 3-class classification of NC vs. MCI vs. AD, were conducted. The results are presented in Table 2.

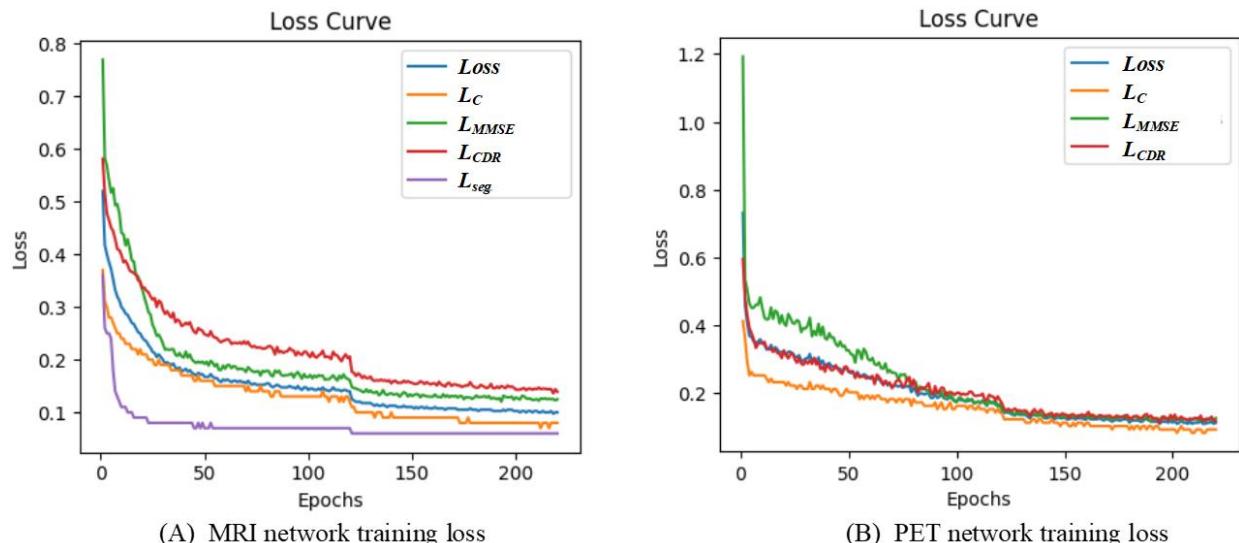


Figure 3. The training loss of MRI and PET network

Table 2. The performance of the proposed approach

Classification	NC F1	MCI F1	AD F1	F1 Average	Accuracy
NC vs. MCI vs. AD	72.7 \pm 0.9%	57.8 \pm 0.9%	78.6 \pm 0.6%	69.7 \pm 0.7%	69.3 \pm 0.7%
NC vs. AD	93.7 \pm 0.3%	-	93.3 \pm 0.3%	93.5 \pm 0.3%	93.5 \pm 0.3%
NC vs. MCI	72.3 \pm 1.1%	74.9 \pm 0.9%	-	73.6 \pm 1%	73.6 \pm 1%
MCI vs. AD	-	80.2 \pm 0.7%	76.8 \pm 0.8%	78.5 \pm 0.7%	78.6 \pm 0.7%

Note: In cells, the two numbers represent the mean and standard deviation of 100 runs. F1 means F1-score

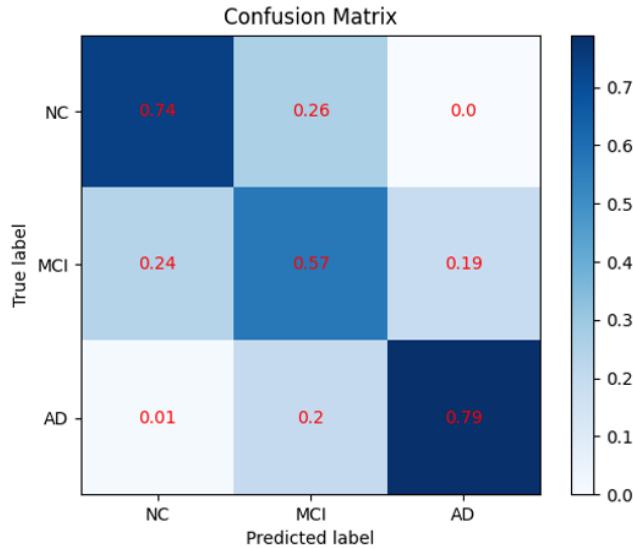


Figure 4. The confusion matrix of 3-class classification

As evident from the aforementioned results, the classification of NC vs. AD is less challenging, resulting in high accuracy. Given that MCI represents an intermediate and progressive state between NC and AD, distinguishing between early MCI and NC, as well as late MCI and AD, is more complex and prone to confusion. So, the accuracy of classification for NC vs. MCI and MCI vs. AD is not as well as for NC vs. AD. For triple categorization, the accuracy drops even further. In order to show more clearly the categorization between the different categories, the confusion matrix of the 3-class classification is shown in Figure 4, from which, it can

be found that the misdiagnoses of NC and AD were almost zero, all the misdiagnosed happened between NC and MCI or between MCI and AD.

4.3 Ablation study

To reveal the contributions of different modalities and components in the proposed approach, experiments were conducted in different settings, and the results are listed in Table 3. Row #1 means the complete data and steps of the approach, which has the best performance. Row #2 replaces the ELM classifier with the SVM classifier, and the accuracy drops by 1.3%. Row #3 indicates the contribution of PCA, the accuracy would decrease 2.1% without PCA. Rows #4 and #5 reveal the importance of the multi-task strategy, especially the segmentation task. Without segmentation task, the accuracy decreased by 3.8%, and without all multitasking, the decrease would be 4.2%. Rows #6 to #8 demonstrate the role of different modality data. Without MRI data, the performance suffered a great decrease, the accuracy dropped by 4.7%, and the other data has a similar contribution as MRI. The PET data has less contribution compared with MRI and other data, for the accuracy decrease is 1.5% with the absence of PET data. Rows #9 and #10 demonstrate the performance of only one modality adopted, and the performance is relatively low, which indicates the importance of the multi-modal method. We used the independent samples t-test to assess the reliability of the mean differences between the proposed method and other methods, and the p-values are listed in the final column of Table 3. These p-values ($p\text{-value} < 0.05$) show that the proposed approach is highly statistically significant when compared to other approaches.

Table 3. The ablation studies of the proposed approach (3-Class classification of NC VS. MCI VS. AD)

Row	MRI Data	PET Data	Other Data	MMSE & CDR Task	Segmentation Task	PCA Classifier	F1 Score	Accuracy	P-Values
#1	√		√	√	√	√	69.7±0.7%	69.3±0.7%	-
#2	√	√	√	√	√	√	68.3±0.7%	68±0.7%	0.001
#3	√	√	√	√	√	√	67.6±0.6%	67.2±0.6%	<0.001
#4	√	√	√	√		√	65.9±0.5%	65.5±0.5%	<0.001
#5	√	√	√			√	65.5±0.9%	65.1±0.9%	<0.001
#6	√	√		√	√	√	65.3±0.6%	64.7±0.6%	<0.001
#7	√		√	√	√	√	68.3±0.7%	67.8±0.7%	<0.001
#8		√	√	√		√	65.2±0.7%	64.6±0.7%	<0.001
#9	√			√	√	√	62.0±0.5%	61.4±0.5%	<0.001
#10		√		√	√	√	60.9±0.7%	60.5±0.7%	<0.001

Note: The best results are highlighted in bold

Based on the results presented in Table 3, we can find that both MRI and PET neuroimaging modalities are effective data sources for AD diagnosis. However, using a single neuroimaging modality alone does not yield optimal performance, combining both MRI and PET can improve diagnostic accuracy. Non-neuroimaging data, such as CSF biomarkers, genetic information, and demographic data, also serve as crucial features for AD diagnosis. Notably, when MRI is combined with these non-imaging features, the diagnostic accuracy approaches that of the method proposed in this paper, highlighting their significant contributions. The hippocampus regions typically undergo atrophy as AD progresses, making their structural morphology and volume critical diagnostic features. Our proposed method explicitly leverages this by incorporating hippocampus segmentation as one of the neural network's multi-tasks. This not only enhances the effectiveness of neural network training but also implicitly

guides the model to focus on this clinically relevant region. Furthermore, we use hippocampal volume as one of the features for AD diagnosis, which effectively boosts diagnostic accuracy. This directly demonstrates the importance of the hippocampus region.

4.4 Sensitivity analysis of multitask loss weights

To further address your point regarding the robustness of our approach and the impact of different multitask weighting coefficients, we have conducted a comprehensive sensitivity analysis on the weighting coefficients. For each weighting coefficient, its value was systematically varied using powers of two (e.g., $2^{-2}, 2^{-1}, 2^0, 2^1, 2^2$) while keeping other coefficients at their baseline values. This allows isolating the impact of each coefficient. The results of the model's performance of each combination of weights are showed in Figure 5. The

results of the model's performance under varying loss function weights reveal differential sensitivity. When w_c was reduced, accuracy significantly dropped to 65.7%, indicating the classification task's critical importance. Conversely, increasing w_c showed less severe degradation. Similarly, perturbations to w_{mmse} , w_{cdr} and w_{seg} exhibited remarkable

robustness. Performance remained within a narrow range, with maximum drops from the peak being only 2.2 percentage points, respectively. This suggests the model is resilient to variations in regression and segmentation task emphasis, but more sensitive to the classification objective.

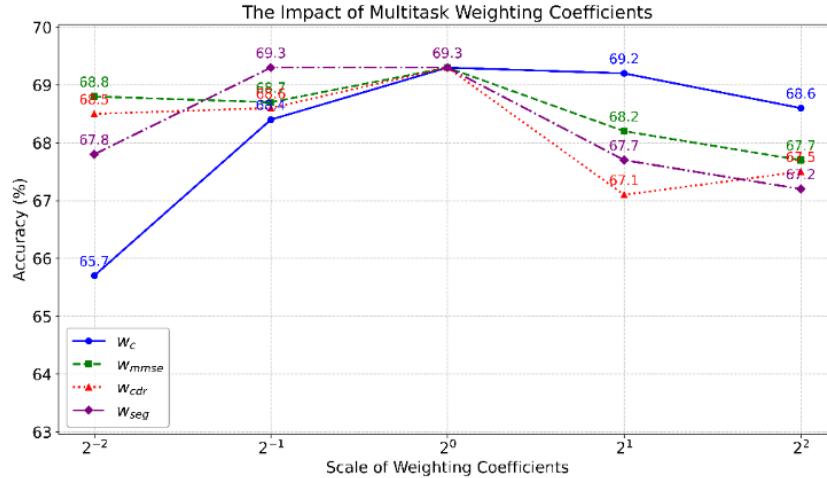


Figure 5. The impact of multitask weighting coefficients

Table 4. Comparison with previous methods

Studies	Methods	F1 Score	Accuracy
[12]	Asymmetry enhanced attention network	-	62.7%
[36]	Gaussian discriminative component analysis	-	67.7%
[37]	Multi-atlases multi-layer perceptron approach	68%	67%
[21]	Modified Tresnet neural network	-	61.8%
[38]	Multi-diagnostic and generalizable approach	-	62.1%
[13]	Multimodal cross-attention AD diagnosis framework	61.85%	64.03%
[39]	Pearson's correlation and empirical cumulative distribution	-	65.46%
[40]	Pearson's correlation and gradient boosting classifier	66.32%	68.2%
[41]	Hybrid region and population hypergraph neural network	55.64%	59.95%
This study	CSMT-Net based multimodal approach	69.7%	69.3%

Note: The best results are highlighted in bold

4.5 Comparison with other methods

We have also evaluated the proposed approach against prior 3-class multistage AD diagnosis studies. The comparison results are listed in Table 4. From these results, it can be found that the proposed method surpasses other methods in both F1 score and accuracy. The F1 score and accuracy of the proposed method are higher than previous methods by at least 1.7% and 1.6%. These results demonstrate the promising performances of the proposed method compared to other methods.

5. DISCUSSION

In this paper, a multitask deep neural network named CSMT-Net is proposed to extract AD-related features from MRI and PET data. These features are combined with CSF biomarkers, Apoe4 genes, age, gender, and education data to enhance the accuracy of AD multistage diagnosis. A U-net upsampling branch is incorporated into the convolutional neural network backbone framework to perform the hippocampal segmentation task in the MRI feature extraction network. The segmentation task enables the neural network to acquire morphology knowledge about brain tissue structure, enhancing the efficacy of MRI deep features. Furthermore, the

volumes of the hippocampus can also serve as AD-related features, enhancing the accuracy of AD diagnosis. Due to the unavailability of manual segmentation for the hippocampus, the FSL tool was used to generate hippocampus segmentation as the segmentation training labels. The FSL segmentation tool, despite not relying on manual segmentation by an expert, is an algorithm with high segmentation accuracy. It incorporates prior knowledge of hippocampal segmentation, enabling the deep neural network to acquire significant MRI structure knowledge and achieve effective segmentation ability.

Multiclass AD diagnosis is a challenging task; the performance of multiclass diagnosis is significantly lower than that of binary diagnosis for AD vs. NC. This is due to the fact that the conversion of NC to MCI, as well as the conversion of MCI to AD, is a gradual process. In this conversion process, there is no obvious boundary between early MCI and NC or late MCI and AD. Therefore, MCI and NC or MCI and AD are easily confused, and it can also be seen from the confusion matrix of the experimental results that most of the classification errors occur between MCI and NC or MCI and AD. Therefore, it is a difficult task to achieve high-precision AD multi-classification. It can also be seen that the key to improving the accuracy of AD multi-classification is to improve the differentiation between MCI and the other two groups, i.e., the classification of NC and MCI and the

classification of MCI and AD. For the structure of the hippocampal regions gradually changed during AD progression; by introducing the hippocampus segmentation task, the CSMT-Net could learn information about hippocampal alterations, which could indicate the AD progression and be helpful for the multistage diagnoses of AD.

The experiment results in Table 2 and the confusion matrix in Figure 4 indicate that the classification accuracy of NC and MCI is lower than that of MCI and AD. For this reason, we assume that in the stage of NC and early MCI, the degree of lesion is slight, and the degree of some brain tissue changes or biomarker abnormalities are mild, and these mild changes are not easy to distinguish, so the classification accuracy is lower. Whereas, in the stages of late MCI to AD, there are more lesions, these changes will be more obvious with the increase of the disease, so it is relatively easier to distinguish, so the classification accuracy is higher. However, discriminating MCI patients from the NC cohort is more meaningful because AD patients are easier to treat in the early stage, so a high performance of NC and MCI classification plays a greater role in the prevention of AD.

MMSE and CDR are not utilized as features directly in this paper, because of their substantial correlation with the group labels, as they are primarily employed in the clinical diagnosis of AD. Utilizing them as features can significantly enhance classification accuracy, but it may lead to biased and overestimated findings. Thus, in this paper, the scores were just utilized as training labels for the neural network and not as features.

The proposed deep learning model integrates MRI and PET neuroimaging, CSF features, genetic risk factors, and demographic data for the diagnosis of AD and MCI. Diagnosing MCI is particularly vital as it represents an early stage of AD, enabling timely intervention and preventative strategies. To facilitate the integration of the proposed approach into clinical workflows, the acquisition of multimodal data is most important. Data such as MRI scans, genetic risk factors, and demographic information are generally more accessible and less invasive for patients in routine clinical practice. MRI is a standard neuroimaging technique, and genetic testing and demographic data collection are common procedures. Conversely, obtaining PET neuroimaging and CSF features poses greater challenges. PET scans involve radiation exposure and are more costly, while CSF collection via lumbar puncture is an invasive procedure that carries certain risks and patient discomfort. In a practical clinical workflow, a tiered approach could be adopted. Initial screening might primarily utilize the more accessible data (MRI, genetics, demographics). For cases with inconclusive results or higher suspicion, the more invasive but highly informative PET and CSF data could then be considered to confirm diagnosis or assess disease progression.

Although the proposed approach demonstrates strong predictive performance, it has some limitations. As a deep neural network, the proposed model, particularly when integrating diverse multi-modal inputs, can present challenges in direct interpretability. Understanding the exact contribution and interplay of each specific feature to a given prediction can be complex. The strategies for post-hoc interpretability could be explored in future work to gain deeper insights into the model's decision-making process. While ADNI is a high-quality dataset, real-world clinical data often present greater variability and more extensive missingness. The proposed model's reliance on a comprehensive set of multi-modal data

(MRI, PET, CSF, genetic, demographic) means that missing data could impact its applicability outside of well-controlled research settings. Meanwhile, the potential biases inherent in the ADNI cohort, which is predominantly of European descent, limit generalizability to more diverse populations. The proposed model is primarily designed for multi-stage classification based on cross-sectional multi-modal data. While ADNI provides longitudinal data, the proposed approach does not fully leverage the temporal dynamics and progression patterns inherent in these longitudinal measurements to predict disease trajectory or conversion risk over time. The explicit modeling of longitudinal changes could offer a more nuanced understanding of disease progression, and it would be a significant area for our future research and model development.

6. CONCLUSION

In this study, a CSMT-Net-based multimodal approach is developed for AD neuroimaging feature extraction. Specifically, deep features associated with AD are extracted from MRI and PET scans using a well-trained CSMT-Net. These features are then subjected to PCA for dimensionality reduction and combined with additional data, including CSF biomarkers, the APOE4 gene, age, gender, and education. An ELM classifier is utilized for multiclass classification using these processed features. Based on the experimental results, the proposed CSMT-Net significantly improves multiclass diagnosis performance. The neuroimaging feature extracted by the CSMT-Net contains sufficient AD-related information, leading to a notable performance improvement, with an accuracy of 69.3% and an F1-score of 69.7% for AD multiclass diagnosis, surpassing the performance of previous studies such as single-task deep neural networks. To meet the requirements of clinical applications and provide more reliable technical support in the biomedical field, future work may incorporate additional modalities or whole-brain segmentation tasks to further improve diagnosis performance.

ACKNOWLEDGEMENTS

This work was supported by Fujian Provincial Natural Science Foundation of China (Grant No.: 2023I0044).

REFERENCE

- [1] Mank, A., Rijnhart, J.J., van Maurik, I.S., Jönsson, L., et al. (2022). A longitudinal study on quality of life along the spectrum of Alzheimer's disease. *Alzheimer's Research & Therapy*, 14(1): 132. <https://doi.org/10.1186/s13195-022-01075-8>
- [2] DeTure, M.A., Dickson, D.W. (2019). The neuropathological diagnosis of Alzheimer's disease. *Molecular Neurodegeneration*, 14(1): 32. <https://doi.org/10.1186/s13024-019-0333-5>
- [3] Tom, S.E., Hubbard, R.A., Crane, P.K., Haneuse, S.J., et al. (2015). Characterization of dementia and Alzheimer's disease in an older population: Updated incidence and life expectancy with and without dementia. *American Journal of Public Health*, 105(2): 408-413. <https://doi.org/10.2105/AJPH.2014.301935>

[4] Dang, M., Yang, C., Chen, K., Lu, P., Li, H., Zhang, Z. (2023). Hippocampus-centred grey matter covariance networks predict the development and reversion of mild cognitive impairment. *Alzheimer's Research & Therapy*, 15(1): 27. <https://doi.org/10.1186/s13195-023-01167-z>

[5] Rao, B.S., Aparna, M., Kolisetty, S.S., Janapana, H., Koteswararao, Y.V. (2024). Multi-class classification of Alzheimer's disease using deep learning and transfer learning on 3D MRI images. *Traitemet du Signal*, 41(3), 1397-1404. <https://doi.org/10.18280/ts.410328>

[6] Reza, D.S.A.A., Afrin, S., Ullah, M.A., Kha, S.K., Toma, S.C., Roy, R., Ali, L.E. (2023). MR image feature analysis for Alzheimer's disease detection using machine learning approaches. *Information Dynamics and Applications*, 2(3): 143-152. <https://doi.org/10.56578/ida020304>

[7] Levin, F., Ferreira, D., Lange, C., Dyrba, M., et al. (2021). Data-driven FDG-PET subtypes of Alzheimer's disease-related neurodegeneration. *Alzheimer's Research & Therapy*, 13(1): 49. <https://doi.org/10.1186/s13195-021-00785-9>

[8] Tu, Y., Lin, S., Qiao, J., Zhuang, Y., Wang, Z., Wang, D. (2024). Multimodal fusion diagnosis of Alzheimer's disease based on FDG-PET generation. *Biomedical Signal Processing and Control*, 89: 105709. <https://doi.org/10.1016/j.bspc.2023.105709>

[9] Papaliagkas, V., Kalinderi, K., Vareltzis, P., Moraitou, D., Papamitsou, T., Chatzidimitriou, M. (2023). CSF biomarkers in the early diagnosis of mild cognitive impairment and Alzheimer's disease. *International Journal of Molecular Sciences*, 24(10): 8976. <https://doi.org/10.3390/ijms24108976>

[10] Shao, W., Peng, Y., Zu, C., Wang, M., Zhang, D. (2020). Hypergraph based multi-task feature selection for multimodal classification of Alzheimer's disease. *Computerized Medical Imaging and Graphics*, 80: 101663. <https://doi.org/10.1016/j.compmedimag.2019.101663>

[11] Zhang, Y., Wang, S., Xia, K., Jiang, Y., Qian, P. (2021). Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion. *Information Fusion*, 66: 170-183. <https://doi.org/10.1016/j.inffus.2020.09.002>

[12] Wang, C., Wei, Y., Li, J., Li, X., et al. (2022). Asymmetry-enhanced attention network for Alzheimer's diagnosis with structural magnetic resonance imaging. *Computers in Biology and Medicine*, 151: 106282. <https://doi.org/10.1016/j.combiomed.2022.106282>

[13] Aparna, M., Rao, B.S. (2023). Enhanced classification of Alzheimer's disease stages via weighted optimized deep neural networks and MRI image analysis. *Traitemet du Signal*, 40(5): 2215-2223. <https://doi.org/10.18280/ts.400538>

[14] Tong, T., Gray, K., Gao, Q., Chen, L., Rueckert, D. (2017). Multi-modal classification of Alzheimer's disease using nonlinear graph fusion. *Pattern Recognition*, 63: 171-181. <https://doi.org/10.1016/j.patcog.2016.10.009>

[15] El-Sappagh, S., Abuhmed, T., Islam, S.R., Kwak, K.S. (2020). Multimodal multitask deep learning model for Alzheimer's disease progression detection based on time series data. *Neurocomputing*, 412: 197-215. <https://doi.org/10.1016/j.neucom.2020.05.087>

[16] Tabarestani, S., Aghili, M., Eslami, M., Cabrerizo, M., et al. (2020). A distributed multitask multimodal approach for the prediction of Alzheimer's disease in a longitudinal study. *NeuroImage*, 206: 116317. <https://doi.org/10.1016/j.neuroimage.2019.116317>

[17] Maruszak, A., Silajdžić, E., Lee, H., Murphy, T., et al. (2023). Predicting progression to Alzheimer's disease with human hippocampal progenitors exposed to serum. *Brain*, 146(5): 2045-2058. <https://doi.org/10.1093/brain/awac472>

[18] Kim, T.A., Syty, M.D., Wu, K., Ge, S. (2022). Adult hippocampal neurogenesis and its impairment in Alzheimer's disease. *Zoological Research*, 43(3): 481-496. <https://doi.org/10.24272/j.issn.2095-8137.2021.479>

[19] Çelebi, S.B., Emiroğlu, B.G. (2023). A novel deep dense block-based model for detecting Alzheimer's disease. *Applied Sciences-Basel*, 13(15): 8686. <https://doi.org/10.3390/app13158686>

[20] Meng, X., Liu, J., Fan, X., Bian, C., et al. (2022). Multimodal neuroimaging neural network-based feature detection for diagnosis of Alzheimer's disease. *Frontiers in Aging Neuroscience*, 14: 911220. <https://doi.org/10.3389/fnagi.2022.911220>

[21] Xu, Z., Deng, H., Liu, J., Yang, Y. (2021). Diagnosis of Alzheimer's disease based on the modified tresnet. *Electronics*, 10(16): 1908. <https://doi.org/10.3390/electronics10161908>

[22] Park, S., Hong, C.H., Lee, D.G., Park, K., Shin, H., et al. (2023). Prospective classification of Alzheimer's disease conversion from mild cognitive impairment. *Neural Networks*, 164: 335-344. <https://doi.org/10.1016/j.neunet.2023.04.018>

[23] Liu, Z., Maiti, T., Bender, A.R. (2021). A role for prior knowledge in statistical classification of the transition from mild cognitive impairment to Alzheimer's disease. *Journal of Alzheimer's Disease*, 83(4): 1859-1875. <https://doi.org/10.3233/JAD-201398>

[24] Dong, A., Zhang, G., Liu, J., Wei, Z. (2022). Latent feature representation learning for Alzheimer's disease classification. *Computers in Biology and Medicine*, 150: 106116. <https://doi.org/10.1016/j.combiomed.2022.106116>

[25] Feng, J., Zhang, S.W., Chen, L. (2021). Extracting ROI-based contourlet subband energy feature from the sMRI image for Alzheimer's disease classification. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(3): 1627-1639. <https://doi.org/10.1109/TCBB.2021.3051177>

[26] Zhou, K., Piao, S., Liu, X., Luo, X., Chen, H., Xiang, R., Geng, D. (2023). A novel cascade machine learning pipeline for Alzheimer's disease identification and prediction. *Frontiers in Aging Neuroscience*, 14: 1073909. <https://doi.org/10.3389/fnagi.2022.1073909>

[27] Khojaste-Sarakhs, M., Haghghi, S.S., Ghomi, S.F., Marchiori, E. (2022). Deep learning for Alzheimer's disease diagnosis: A survey. *Artificial Intelligence in Medicine*, 130: 102332. <https://doi.org/10.1016/j.artmed.2022.102332>

[28] Pan, D., Luo, G., Zeng, A., Zou, C., et al. (2022). Adaptive 3DCNN-based interpretable ensemble model for early diagnosis of Alzheimer's disease. *IEEE Transactions on Computational Social Systems*, 11(1): 247-266. <https://doi.org/10.1109/TCSS.2022.3223999>

[29] El-Sappagh, S., Saleh, H., Ali, F., Amer, E., Abuhmed, T. (2022). Two-stage deep learning model for

Alzheimer's disease detection and prediction of the mild cognitive impairment time. *Neural Computing and Applications*, 34(17): 14487-14509. <https://doi.org/10.1007/s00521-022-07263-9>

[30] Liu, M., Zhang, J., Adeli, E., Shen, D. (2018). Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis. *IEEE Transactions on Biomedical Engineering*, 66(5): 1195-1206. <https://doi.org/10.1109/TBME.2018.2869989>

[31] Dong, Q., Zhang, J., Li, Q., Wang, J., et al. (2020). Integrating convolutional neural networks and multi-task dictionary learning for cognitive decline prediction with longitudinal images. *Journal of Alzheimer's Disease*, 75(3): 971-992. <https://doi.org/10.3233/JAD-190973>

[32] Dunk, M.M., Driscoll, I. (2022). Total cholesterol and APOE-related risk for Alzheimer's disease in the Alzheimer's disease neuroimaging initiative. *Journal of Alzheimer's Disease*, 85(4): 1519-1528. <https://doi.org/10.3233/JAD-215091>

[33] Gao, R., Yang, F., Yang, W., Liao, Q. (2018). Margin loss: Making faces more separable. *IEEE Signal Processing Letters*, 25(2): 308-312. <https://doi.org/10.1109/LSP.2017.2789251>

[34] Zhan, X., Liu, Y., Cecchi, N.J., Gevaert, O., Zeineh, M.M., Grant, G.A., Camarillo, D.B. (2022). Finding the spatial co-variation of brain deformation with principal component analysis. *IEEE Transactions on Biomedical Engineering*, 69(10): 3205-3215. <https://doi.org/10.1109/TBME.2022.3163230>.

[35] Vásquez-Coronel, J.A., Mora, M., Vilches, K. (2023). A review of multilayer extreme learning machine neural networks. *Artificial Intelligence Review*, 56(11): 13691-13742. <https://doi.org/10.1007/s10462-023-10478-4>

[36] Fang, C., Li, C., Forouzannezhad, P., Cabrerizo, M., et al. (2020). Gaussian discriminative component analysis for early detection of Alzheimer's disease: A supervised dimensionality reduction algorithm. *Journal of Neuroscience Methods*, 344: 108856. <https://doi.org/10.1016/j.jneumeth.2020.108856>

[37] Hong, X., Huang, K., Lin, J., Ye, X., et al. (2022). Combined multi-atlas and multi-layer perception for alzheimer's disease classification. *Frontiers in Aging Neuroscience*, 14: 891433. <https://doi.org/10.3389/fnagi.2022.891433>

[38] Diogo, V.S., Ferreira, H.A., Prata, D. (2022). Early diagnosis of Alzheimer's disease using machine learning: A multi-diagnostic, generalizable approach. *Alzheimer's Research & Therapy*, 14(1): 107. <https://doi.org/10.1186/s13195-022-01047-y>

[39] Mabrouk, B., Hamida, A.B., Mabrouki, N., Bouzidi, N., Mhiri, C. (2024). A novel approach to perform linear discriminant analyses for a 4-way alzheimer's disease diagnosis based on an integration of pearson's correlation coefficients and empirical cumulative distribution function. *Multimedia Tools and Applications*, 83(31): 76687-76703. <https://doi.org/10.1007/s11042-024-18532-1>

[40] Song, J., Huang, H., Liu, J., Wu, J., et al. (2024). Diagnostic potential of eye movements in Alzheimer's disease via a multiclass machine learning model. *Cognitive Computation*, 16(6): 3364-3378. <https://doi.org/10.1007/s12559-024-10346-5>

[41] Wang, J., Wu, L., Kuang, H., Wang, J. (2026). Hybrid region and population hypergraph neural network for mild cognitive impairment detection. *Pattern Recognition*, 169: 111864. <https://doi.org/10.1016/j.patcog.2025.111864>