Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 8, 169-189
2026
Publisher: Learning Gate
DOI: 10.55214/2576-8484.v10i8.13501

Research on computer-aided diagnosis and implementation of Alzheimer's disease based on deep learning

Guangtong Zhou[1*]
[1]School of Sichuan University of Electronic Information, Chengdu, 116025, China; zgt304927736@163.com (G.Z.).

Abstract: MRI and PET are commonly used in the early diagnosis of Alzheimer's disease (AD). MRI shows the brain's structure, and PET shows how the brain works. A deep learning model is built to diagnose AD. CNN can find small details in 3D brain images. A transformer checks the way that different parts of the brain connect. Thus, computers understand the images better. The database is built by using data from the ADNI database. The data include AD, CN, and MCI. The data were split into training, validation, and testing groups for training and improvement after image processing. The proposed model was evaluated on both two-class and three-class tasks. In the two-class task, the model achieved an accuracy of 94.03% on MRI images and 92.54% on PET images. In the three-class task, the corresponding accuracy was 69.16% for MRI and 68.25% for PET. The results indicate that combining CNN with Transformer improves feature representation for brain images and provides better classification performance than the comparison models. In summary, the proposed CNN-Transformer model efficiently integrates the local structural and global contextual information from MRI and PET images. The proposed method can offer a practical reference for the computer-aided diagnosis of Alzheimer's disease.

Keywords: *Alzheimer's disease, Convolutional neural network, Deep learning, Magnetic resonance imaging.*

1. Introduction

Dementia has become an important health problem worldwide, especially in aging societies. Every year, nearly 10 million new cases of dementia are reported. Alzheimer's disease (AD) makes up most of them. AD can affect memory, cognition, and daily living ability. This puts a huge strain on hospitals and families [1].

The development of AD has many causes, such as genetic and environmental conditions. At present, amyloid-beta (Aβ) deposition and hyperphosphorylated Tau protein accumulation are regarded as two major pathological characteristics of the disease. However, the underlying mechanisms have not yet been fully clarified. As a result, effective disease-modifying treatments remain unavailable.

Current treatment mainly aims to delay disease progression and relieve symptoms. Usually, drug therapy improves the functioning of neurotransmitters, while cognitive training and psychological interventions are often combined to maintain cognitive function and improve quality of life. Moreover, research demonstrates that adopting healthy lifestyle behaviors can potentially decrease the likelihood of developing AD or delay its development, such as regular physical activity, a nutritious diet, and active social engagement.

Researchers have improved understanding of AD through advances in medical imaging. Imaging techniques are important tools to study AD. Among these technologies, MRI and PET are used in clinical research. MRI can reflect changes in the brain. PET can provide information about brain metabolism and pathological changes. These imaging techniques provide important references for early AD detection and disease evaluation.

History: Received: 27 May 2026; Revised: 30 July 2026; Accepted: 27 August 2026; Published: 24 August 2026
* Correspondence: zgt304927736@163.com

Population aging has become prominent in China. The number of people affected by Alzheimer's disease has also continued to rise. AD usually develops slowly, and obvious symptoms may not appear during the early stage.

Recently, AI has been widely used in medical image analysis. It can automatically learn useful features from MRI and PET images. It can also identify subtle changes in brain images that may be difficult to detect through manual observation. Therefore, deep learning-based CAD systems can provide objective information for AD diagnosis and assist clinical decision-making.

AD is still a frequent neurodegenerative disease among the elderly. Current treatments mainly aim to delay disease progression rather than achieve complete recovery. Therefore, detection is particularly significant for disease management. With medical imaging and artificial intelligence developing, MRI- and PET-based CAD methods may provide effective approaches for early detection.

2. Related Work

The accurate diagnosis of AD remains challenging owing to the complexity of AD pathology and its slow progression. Current studies are still limited in effective neuroimaging feature extraction, fusion of local and global information, and improvement of model adaptability. Hence, this section reviews the theoretical basis of AD imaging techniques and deep learning methods. The development of current approaches is summarized, and the limitations of existing studies are discussed to provide a basis for the design of the proposed model.

2.1. Theoretical Foundation

AD is a disease of the brain. It leads to a decrease in cognitive skills over time. The main problems in the brain are the buildup of beta-amyloid, abnormal tau proteins, and the death of nerve cells. These changes can start years before symptoms show up.

Doctors use MRI and PET scans to diagnose AD. MRI has high clarity. The brain structure is clearly shown. For example, patients with AD have a degenerated hippocampus and a thinner cortex. Advanced MRI methods such as functional MRI and diffusion tensor imaging also show how the brain regions connect and study white matter. But MRI only shows anatomy. It can't see molecular changes directly. PET is useful for diagnosing AD, staging the disease, and checking treatments.

AI is growing fast. Deep learning is now a key tool for analyzing medical images. Unlike the old methods, deep learning models learn complex patterns in medical images. Deep learning models are good at picking out spatial features from images, so they are widely used to help diagnose AD.

2.2. Literature Review

Lately, researchers use deep learning models to help diagnose Alzheimer's. There are three things: extracting features from medical images, designing the models, and improving the classification. Even though current methods work well, there are some difficulties. Local and global features don't mix enough, spatial information isn't used fully, and the models can't always work well on different datasets. Therefore, reviewing current research progress is necessary for understanding the development of AD computer-aided diagnosis methods and designing improved models.

2.2.1. CNN-Based CAD for AD

CNN-based methods are used to classify AD images because of their ability to learn spatial features from medical images. Various CNN architectures and fusion strategies have been proposed to improve feature representation. Lu et al. [2] presented a multi-scale deep neural network to fuse MRI and FDG-PET images for extracting features at different scales and applied it for early AD diagnosis. Lin et al. [3] built a 3D reversible GAN to construct the relationship between MRI and PET images and enhance classification performance through early feature fusion. Song et al. [4] proposed a multimodal fusion approach that fuses gray matter information, thereby increasing the exploitation of complementary information between modalities. Liu et al. [5] proposed a cascaded CNN network in which multiple CNNs

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 8: 169-189, 2026
DOI: 10.55214/2576-8484.v10i8.13501

were used to extract volumetric features, which were then fused by a 2D CNN. Feng et al. [6] used a 3D CNN and BiLSTM to fuse MRI and PET features at a late stage, which improved the ability to model spatial and sequential information. Huang et al. [7] also investigated automatic AD classification using multimodal 3D CNN architectures and studied different fusion strategies.

However, there are some studies reporting that single-modality imaging methods can achieve similar or even better performance than some multimodal fusion methods. Narazani et al. [8] reviewed the existing 3D CNN-based MRI and PET fusion methods and concluded that these methods did not always outperform the models with only PET. This finding suggests that multimodal fusion still faces several challenges, including redundant information between modalities, insufficient interaction between different feature representations, and differences in data distributions across imaging modalities.

2.2.2. Transformer-Based CAD

Transformer-based methods have gradually been introduced into medical image analysis. Transformers can capture relationships between distant image regions and provide new solutions for feature learning and multimodal information integration because of self-attention. In the field of AD diagnosis, Li et al. [9] built DFENet. It combines CNN and Transformers through a dual-branch feature enhancement structure for image fusion across several modalities. Zhang et al. [10] built a multilevel feature-fused, end-to-end CNN network. Gao et al. [11] proposed Mul-T, which uses DenseNet and spatial attention mechanisms to extract features, while a cross-modal Transformer integrates T1-MRI, T2-MRI, and PET information. Miao et al. [12] built a multimodal Transformer fusion network by combining CNN residual modules with Transformer structures for joint feature learning. Tang et al. [13] extracted MRI and PET features using 3D CNN and applied an improved Transformer to explore relationships among different features, improving AD classification performance.

Although these approaches achieved great results, most existing CNN-Transformer frameworks still adopt a sequential structure, where CNNs are mainly responsible for extracting features, with Transformers used for subsequent feature fusion. In such designs, local feature learning and global relationship modeling are performed separately, which may limit the interaction between different levels of information.

2.2.3. Transfer Learning and Residual CNN in CAD for AD

CNN-based methods remain widely used in AD image analysis because of their capacity for spatial feature extraction from 3D images. Following the successful application of U-Net in medical image segmentation, related CNN-based structures have also been explored for AD auxiliary diagnosis. Kavitha et al. [14] presented a U-Net multi-instance learning network for identifying regions of interest from 3D PET images and performing multi-class AD classification. Bi et al. [15] built an unsupervised deep learning framework that uses CNNs to learn imaging features without manual annotations, reducing the demand for labeled data.

When it comes to MRI-based AD analysis, Jia et al. [16] used classification measures including image transformation and feature fusion. They extract fMRI features using an improved 3D PCANet and combine multiple features through Canonical Correlation Analysis. Sharma et al. [17] proposed FDN-ADNet, which extracts features from sagittal MRI slices and combines them with Fuzzy Least Squares Twin Support Vector Machine for AD prediction in order to improve the modeling of 3D spatial information in MRI. For PET imaging, Rogeau et al. [18] devised a VGG16-like network to distinguish 18F-FDG PET into AD, Frontotemporal Dementia, and Healthy Control groups. Qiu et al. [19] proposed a 3D multimodal fusion framework. It had a disease-induced joint studying module, which enhances the representation of important brain regions. It worked well for early AD diagnosis.

Given scarcities and high acquisition costs of medical imaging data, transfer learning has become a critical research vector. For early AD diagnosis, Cheng et al. [20] and Kumar and Nandhini [21] proposed a multi-domain transfer learning network that improved detection via multi-domain feature selection Cheng et al. [20]. Kumar and Nandhini [21] developed FEESCTL, using entropy slicing to

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 8: 169-189, 2026
DOI: 10.55214/2576-8484.v10i8.13501

isolate key MRI regions, followed by VGG-16 transfer learning. Naz et al. [22] leveraged frozen-feature VGG networks for multiclass AD tests. Mehmood et al. [23] proposed layer-wise transfer learning methods combined with brain tissue segmentation. Furthermore, ResNet architectures have seen extensive use due to the advantages of residual connections. Odusami et al. [24], Buvaneswari and Gayathri [25], and Odusami et al. [26] use ResNet18 to build deep learning models. In another study, Ashraf et al. [27] compared 13 pre-trained CNN models, and DenseNet obtained better classification results to distinguish AD, MCI, and control groups.

Combinations of transfer learning and Transformer-based networks have received increasing attention in AD diagnosis research. Transfer learning can reduce the difficulty of model training when medical imaging datasets are limited, while Transformer structures provide stronger capability for modeling global feature relationships. In order to solve the problem of imbalance, Oktavian et al. [28] built a transfer learning network using ResNet18 and weighted loss functions. Gharaibeh et al. [29] proposed an AD detection model by combining Swin Transformer with multi-scale Feature Pyramid Networks to improve the utilization of multi-scale features. Leela et al. [30] proposed a framework named HEMRDTL for multimodal learning that integrates EEG and CT/MRI data and explores the use of transfer learning for heterogeneous medical data analysis.

Mahmud Joy et al. [31] proposed ViTAD, utilizing an improved Vision Transformer (ViT) to achieve five-stage AD classification on brain MRI. Turrisi et al. [32] proposed a 3D transfer learning adaptation strategy allowing 2D pre-trained models to effectively process 3D medical images.

MCI represents the vital prodromal AD stage, so predicting MCI-to-AD conversion is a primary focus. Li et al. [33] developed a hippocampal MRI model to predict MCI conversion, integrating clinical variables to boost screening performance. Lee et al. [34] proposed an interpretable model fusing voxel-, region-, and patch-level MRI features to map brain abnormalities while enhancing performance. Li et al. [35] proposed SSCHM, a PET classification framework designed for AD and conversion prediction. SSCHM employs stacked LMamba parts for wide-range dependency modeling in 3D voxel sequences, incorporates a layer-wise cross-scale channel attention module for adaptive context fusion, and uses a channel-spatial awareness module with large-kernel convolutions to refine lesion localization. Xiao et al. [36] introduced an improved ResNet50 model, which was called ResNet50-DW-SE, to classify AD. Depthwise separable convolutions replaced traditional convolution operations. This decreases model parameters and improves computational efficiency.

2.3. Synthesis and Discussion

Deep learning works well in diagnosing Alzheimer's. Transfer learning has improved model training when only limited medical data are available. Despite these advances, several problems still need further investigation.

First, many methods adopt a two-stage framework. CNNs are used for feature extraction, and Transformers are responsible for feature fusion. Since these two processes are performed separately, the interaction between local feature learning and global information modeling is limited. As a result, the advantages of Transformers are not fully utilized.

Second, MRI and PET provide different types of information and have different imaging characteristics, including spatial resolution, pathological representation, and signal distribution. Simply combining features from the two modalities may introduce redundant information or increase inconsistency between modalities, which can reduce classification performance.

Third, the amount of available medical imaging data is still limited. Although transfer learning can improve model performance to some extent, most pre-trained models are developed using natural image datasets such as ImageNet. The difference between natural images and 3D medical images may reduce the effect of feature transfer.

Finally, many studies focus on classification between AD and CN. Comparatively fewer studies investigate MCI or the progression from MCI to AD. MCI changes are very small, so the model needs to be really good at spotting the key details in the images.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 8: 169-189, 2026
DOI: 10.55214/2576-8484.v10i8.13501

3. Methodology and Dataset Construction

CNNs and Transformers have different tasks. CNNs look at small and local parts of an image. Transformers use self-attention to see connections between different areas. MRI and PET images contain both local anatomical information and global interactions between brain regions. This study combines 3D CNN and Transformer structures to build a deep learning model.

3D CNN modules are first used to learn local representations from volumetric MRI and PET images. The extracted features are transformed into sequential representations. They are input into the Transformer module for global dependency modeling. Local spatial information and global contextual features are jointly considered during classification.

3.1. Proposed Methodology

A traditional CNN usually consists of convolutional layers, pooling layers, and fully connected layers. The convolution operation uses learnable kernels applied to the input data, enabling the network to learn different feature levels from the image. In the early convolution layers, low-level patterns are mainly learned. More abstract semantic information is gradually learned in the deeper layers.

The mathematical formulation of the 2D convolution operation (as shown in Figure 1) can be expressed as:

$$(I * K)(\mathrm{i},\ \mathrm{j}) = \sum_{\mathrm{m}} \sum_{\mathrm{n}} I(i+m, j+n) K(m,n) \qquad (1)$$

where I denotes the input image, and K denotes the convolution kernel. After convolution, nonlinear activation functions such as ReLU are used to improve the nonlinear fitting ability of the network:

$$\mathrm{f}(x) = \max(0, x) \qquad (2)$$

At the last stage of a CNN, the extracted feature maps are flattened and fed into fully connected layers for classification:

$$\mathrm{z} = WX + b, y = \phi(\mathrm{z}) \qquad (3)$$

where W represents the weight matrix, b denotes the bias term, and ϕ represents the activation function.

For three-dimensional medical images, 3D CNN extends convolution operations from two spatial dimensions to three dimensions, i.e., width, height, and depth. Compared to 2D CNN, this structure can better preserve the continuity of volumetric information and retain the spatial relationships between different brain regions. Therefore, 3D CNN is more suitable for analyzing medical images with three-dimensional structures.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 8: 169-189, 2026
DOI: 10.55214/2576-8484.v10i8.13501

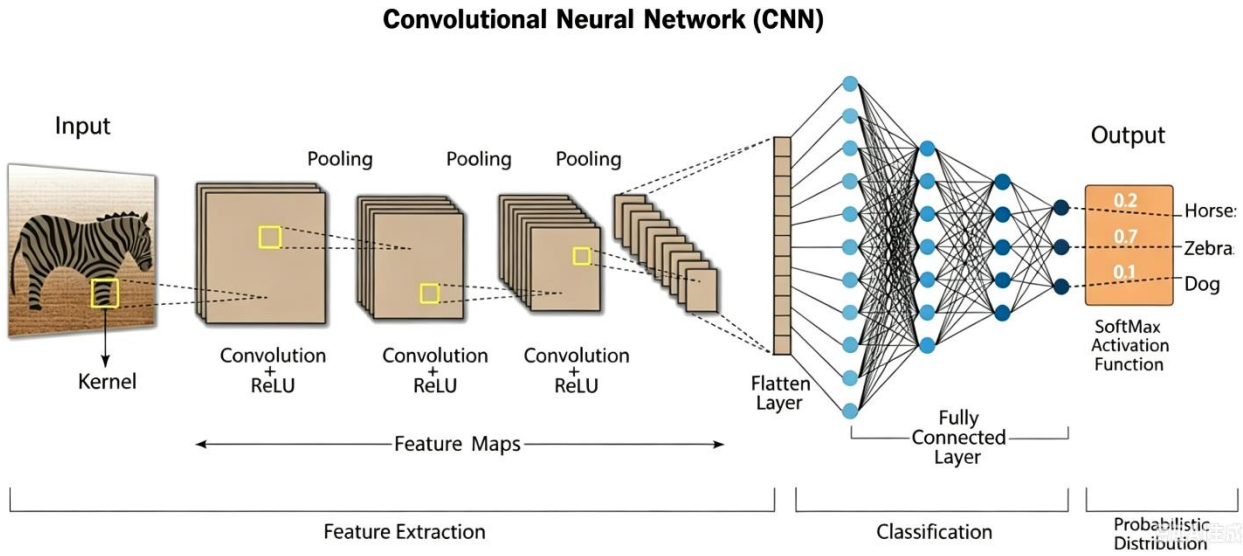

Figure 1.
Structural diagram of a standard CNN model.

Rather than relying on recurrent operations or convolution, the Transformer (Figure 2) includes a self-attention mechanism that can model relationships among different elements in a series.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 8: 169-189, 2026
DOI: 10.55214/2576-8484.v10i8.13501

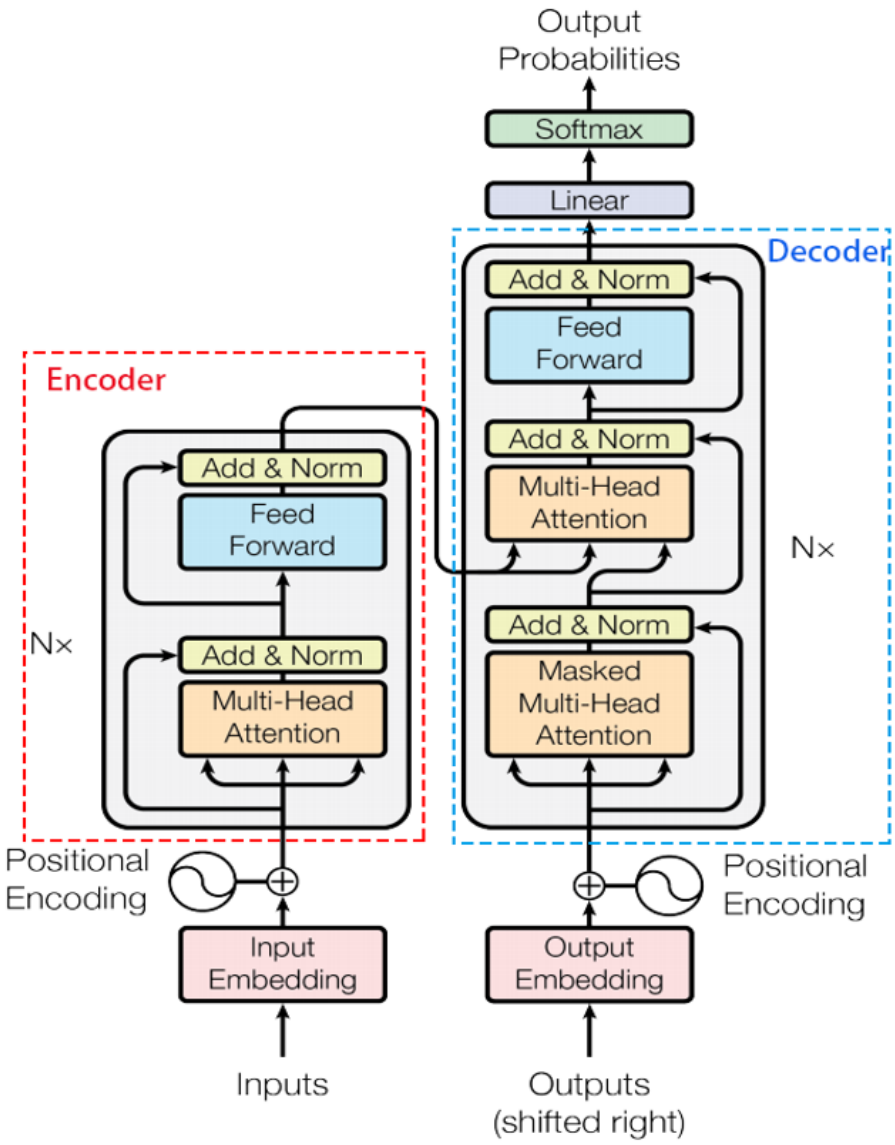

Figure 2.
Structural diagram of the Transformer model.

Scaled Dot-Product Attention is the core building block of Transformers, where the input features are projected into three matrices: Query (Q), Key (K), and Value (V).

$$Attention(Q,K,V) = soft\max\left(\frac{QK^{T}}{\sqrt{d_k}}\right)V \tag{4}$$

To capture information from different feature subspaces, the Transformer further introduces the Multi-Head Attention mechanism:

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 8: 169-189, 2026
DOI: 10.55214/2576-8484.v10i8.13501

$$Multihead(Q,K,V) = Concat(head_1,...,\text{head}_h)W^O \tag{5}$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \tag{6}$$

To inject positional order into token sequences, positional encodings (PE) are added:

$$PE(pos,2i) = \sin(\frac{pos}{10000^{2i/d}}), PE(pos,2i+1) = \cos(\frac{pos}{10000^{2i/d}}) \tag{7}$$

Each Transformer layer also includes a Feed-Forward Network:

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \tag{8}$$

Residual connections coupled with Layer Normalization (LN) ensure training stability:

$$y = \text{LayerNorm}(x + \text{Sublayer}(x)) \tag{9}$$

As deep learning develops, Transformer-based methods are gradually being applied to image analysis. Recent studies have demonstrated their potential for processing MRI and PET images for AD diagnosis.

Preliminary model structures are shown in Figure 3. The network mainly consists of four parts: a 3D CNN local feature extraction module; a feature token generation module; a Transformer-based global feature encoding module; classification module.

The input of the network consists of preprocessed 3D MRI or PET brain images with a size of 1×128×128×128. Here, 1 represents the single-channel input, and 128×128×128 represents the normalized voxel dimensions. Since MRI and PET data contain three-dimensional spatial structures, 3D convolution operations are adopted to extract features while maintaining spatial continuity between adjacent brain regions.

For local feature extraction, the proposed model employs three sequential 3D CNN blocks. Each block consists of a 3D convolutional layer, a ReLU activation function, and a max-pooling layer. The first convolutional block tries to learn the basic spatial patterns of the input images. The following layers gradually learn more complex representations related to brain structure and metabolic changes.

Three pooling operations are adopted to decrease the spatial resolution of feature maps during feature extraction. This process reduces computational requirements while retaining important image information. The CNN module learns hierarchical representations of MRI and PET images through stacked convolutional operations for global feature learning in the Transformer module.

Before global feature learning, an adaptive average pooling layer can decrease the convolutional feature maps to a smaller spatial size. This measure not only decreases the volume of data that the subsequent Transformer needs to process but also enhances model robustness against varying input scales. The compressed 3D feature maps are then converted into a token sequence format. Next, a linear transformation projects each token into a 128-dimensional embedding space for input into the Transformer encoder.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 8: 169-189, 2026
DOI: 10.55214/2576-8484.v10i8.13501

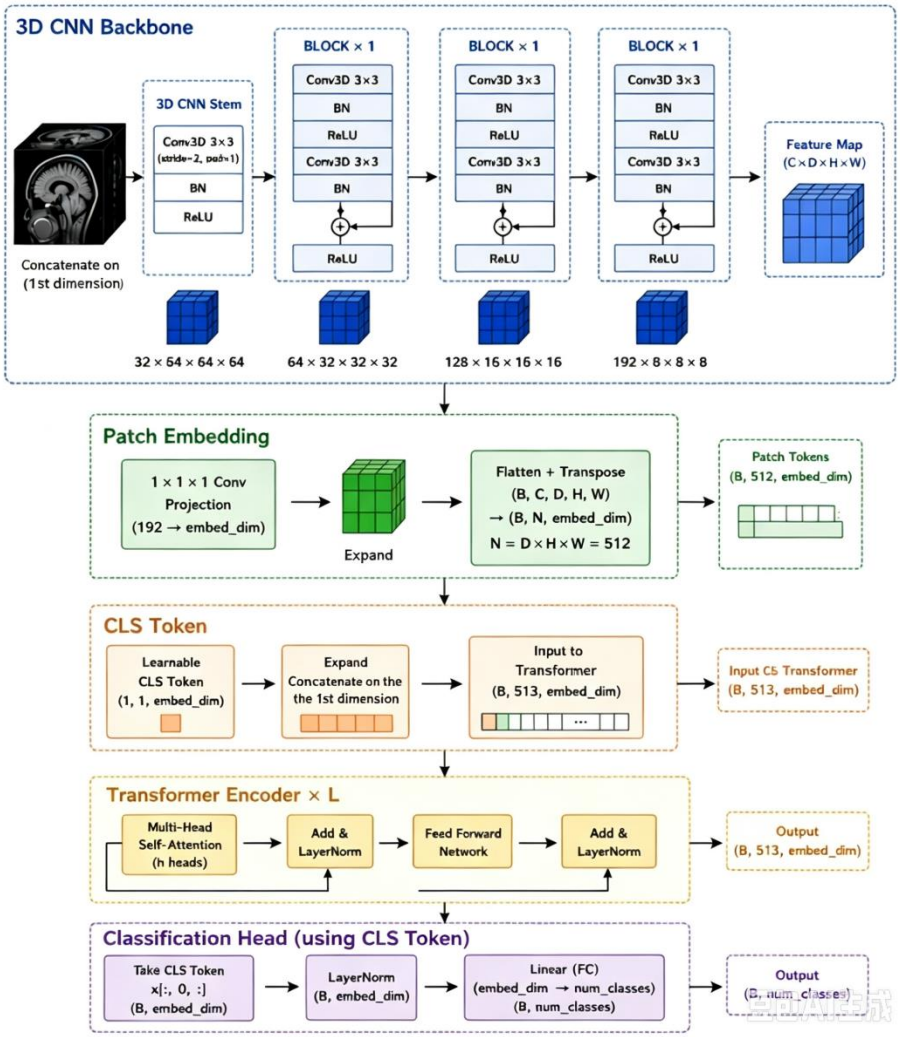

Figure 3.
Preliminary model architecture.

To enable the Transformer to better capture relative positional relationships among sequence tokens, a learnable special token, the [CLS] token, is prepended to the token sequence. The [CLS] token is used to aggregate information across the entire sequence, and its output representation at the final layer is utilized for classification.

During global feature learning, two Transformer Encoders are used to obtain the token sequence. Each Transformer Encoder includes multi-head self-attention mechanisms, an FFN, residual connections, and layer normalization. Specifically, multiple attention heads learn distinct aspects of inter-token interactions, establishing long-range dependencies across distant brain regions. The FFN enhances the ability of the nonlinear model, while residual connections and layer normalization stabilize training and accelerate convergence. The Transformer module is used to model relationships between different brain regions. Through the self-attention mechanism, it can capture connections between distant regions and provide additional global information for MRI and PET image classification.

After feature encoding, the output representation from the Transformer is used for classification. In the baseline model, the [CLS] token is selected as the global feature representation and is subsequently

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 8: 169-189, 2026
DOI: 10.55214/2576-8484.v10i8.13501

passed to a fully connected classifier. The classification layer adopts a linear mapping operation, which converts the 128-dimensional feature vector into a 2-class or 3-class prediction result.

Although the baseline network can extract useful image features, several limitations remain. The CNN module mainly focuses on local feature extraction, while its ability to model global relationships between brain regions is limited. Meanwhile, the original Transformer structure does not contain explicit spatial position information, which may affect its understanding of three-dimensional medical images. In addition, using only the [CLS] token for feature aggregation may not fully represent important regions in MRI and PET images.

To overcome these limitations, the improved model (Figure 4) retains the basic CNN-Transformer framework and focuses on three aspects: enhancement of local feature extraction, improvement of spatial information modeling, and optimization of global feature aggregation.

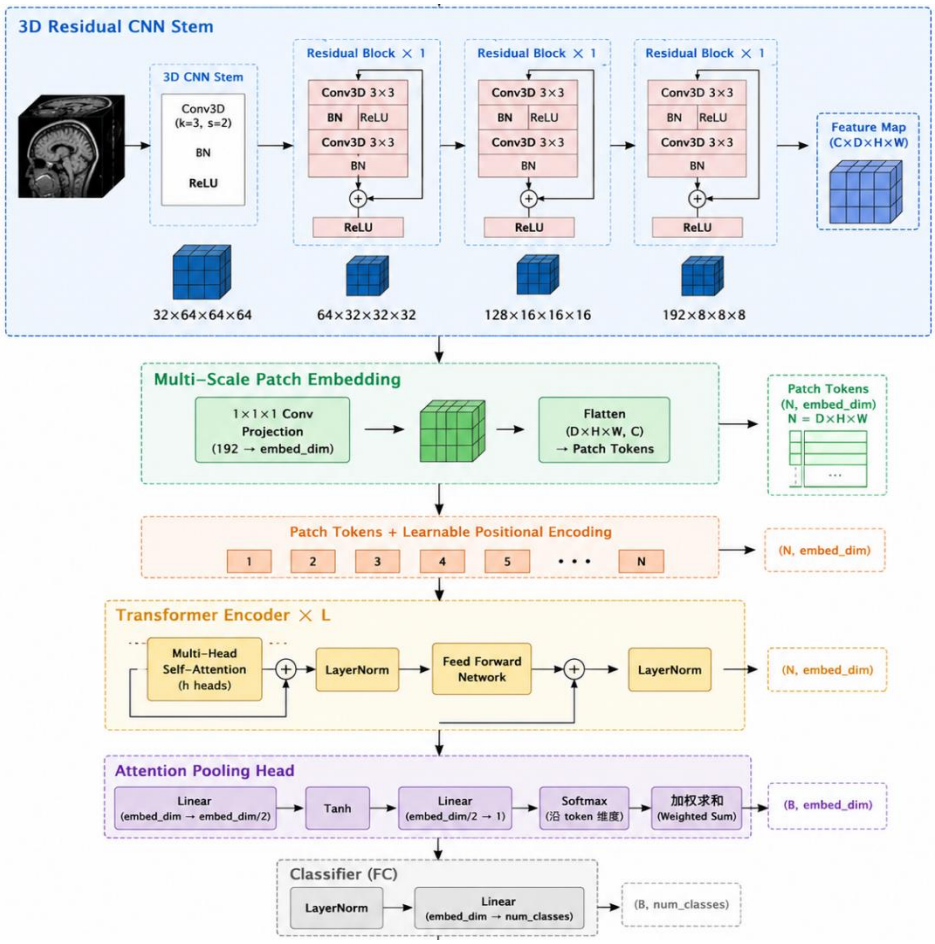

Figure 4.
Improved model architecture.

MRI and PET scans are typical 3D volumetric data in which local spatial relationships are critically important. Consequently, as a first step, a 3D residual convolutional neural network (3D ResCNN) is adopted to learn local anatomical structures, capturing local textures and boundaries, thereby learning differences across brain regions while preserving 3D spatial context. Incorporating residual connections into the 3D CNN architecture allows the network to go deeper without suffering from gradient vanishing or network degradation, as residual connections establish identity mappings that enable shallow inputs to flow directly to deep outputs. Given that MRI and PET are high-dimensional, complex 3D medical

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 8: 169-189, 2026
DOI: 10.55214/2576-8484.v10i8.13501

images, residual blocks not only enhance the representation of local anatomical details but also increase sensitivity to subtle pathological changes.

Because the baseline model generates a small number of tokens, its sequence representation capacity is limited; furthermore, directly flattening the feature maps leads to a severe loss of spatial relationships, leaving the Transformer unable to discern which tokens belong to the same brain region. The 3D Patch Embedding mechanism increases the total token count, allowing the Transformer to observe more brain region details and finer-grained spatial structures. Meanwhile, since complex degenerative relationships exist among different brain regions, tokenizing patches enables the Transformer to effectively learn long-range dependencies across distant brain regions.

The proposed model also modified the multi-head attention mechanisms. We increased the number of heads from 4 to 6 so that the Transformer could learn feature information from more representational subspaces. The more attention heads, the more different types of brain imaging characteristics could be captured simultaneously by the model, including hippocampal atrophy, ventricular enlargement, cortical thickness, gray matter distribution, and other changes related to neurodegeneration. This modification improves the network's capacity to represent complex brain features.

Another improvement is Learnable Positional Encoding. The CNN extracts local features of anatomy. Then the Transformer is applied to model the relations of different brain regions. However, the self-attention mechanism itself does not have positional information, so it does not know the original spatial location of each patch of the image. For MRI and PET images, spatial position is closely related to brain anatomy, and disease-related abnormalities often appear in specific regions. If positional information is not considered, the model may have difficulty distinguishing features from different anatomical locations. To overcome this limitation, learnable positional encoding is introduced to provide each image patch with spatial information. This helps the Transformer preserve the 3D structures of medical images and improves its ability to model spatial relationships.

The feature aggregation strategy was also modified. Conventional Transformer models usually use the [CLS] token as the global representation for image classification. However, this approach is not always suitable for three-dimensional medical images. MRI and PET volumes contain a large number of image patches, while disease-related regions occupy only a small part of the whole brain. As a result, the information contained in a single [CLS] token may not fully represent the important pathological features. In this study, Attention Pooling replaces the traditional [CLS] token. Instead of relying on a single feature vector, the network assigns different weights to all tokens and performs weighted feature aggregation. In this way, more attention is given to image patches containing important brain structures, while the influence of background information is reduced. This leads to a more stable feature representation and improves the final classification performance.

3.2. Dataset Construction and Selection

Because medical imaging data is difficult to obtain, this experiment's dataset was built using publicly available ADNI data, and the data were preprocessed to meet the model training requirements.

3.2.1. ADNI Database

Data were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). Launched in 2004 under the leadership of Lin [37], ADNI is a multicenter collaborative public-private partnership funded by the NIA, NIBIB, the Alzheimer's Association, and private industry. ADNI provides clinical, imaging, genetic, and biochemical biomarker data across multiple study phases.

3.2.2. Cohort Selection

Data criteria were configured via the ADNI Image Collection interface. To ensure sufficient representation of early-stage cognitive decline, the ADNI GO phase was explicitly selected (Figure 5), ensuring adequate Early MCI (EMCI) samples.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 8: 169-189, 2026
DOI: 10.55214/2576-8484.v10i8.13501

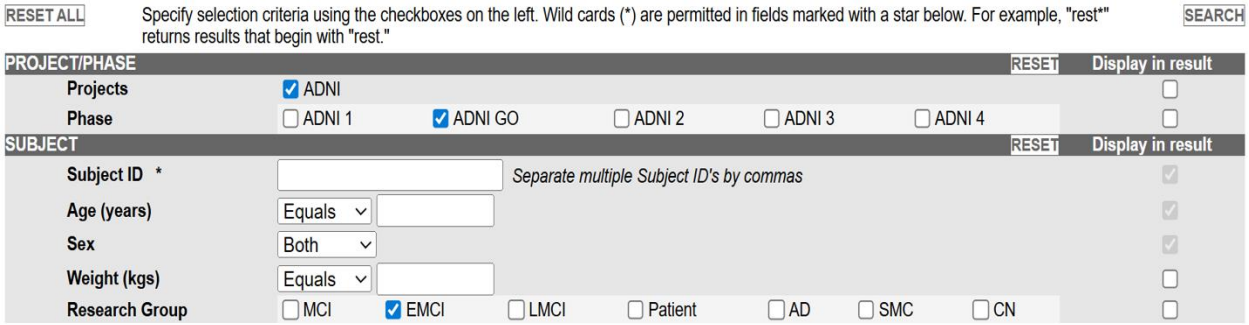

Figure 5.
ADNI IDA phase selection screen.

For structural MRI, T1-weighted scans (MPRAGE and sagittal IR-SPGR sequences) were selected (Figure 6) because of their high contrast between gray and white matter and sensitivity to structural atrophy. FDG-PET scans were retrieved using corresponding multiclass selection rules.

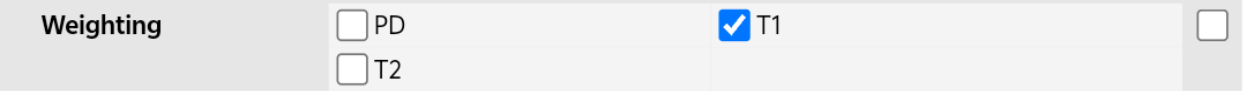

Figure 6.
ADNI IDA weighting selection screen.

The final dataset was partitioned into training, validation, and test groups using a 70%:15%:15% split ratio. The MRI dataset contained 800 total scans (213 AD, 234 CN, 353 MCI). The PET dataset contained 866 total scans (219 AD, 228 CN, 419 MCI). Detailed distributions are listed in Tables 1 and 2.

Table 1.
MRI Dataset Composition.

Dataset Split	Category	Sample Size
Training Set	AD	149
	CN	164
	MCI	247
Validation Set	AD	32
	CN	35
	MCI	53
Test Set	AD	32
	CN	35
	MCI	53

Table 2.
PET Dataset Composition.

Dataset Split	Category	Sample Size
Training Set	AD	147
	CN	166
	MCI	301
Validation Set	AD	36
	CN	31
	MCI	59
Test Set	AD	36
	CN	31
	MCI	59

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 8: 169-189, 2026
DOI: 10.55214/2576-8484.v10i8.13501

3.2.3. Data Preprocessing

Raw DICOM files were converted to NIfTI format using dcm2niix. To prevent the network from learning non-brain features (such as skull contours and scanning boundaries), automated skull-stripping was executed using HD-BET (Figure 7). All steps included:

Background cropping: Bounding-box extraction around non-zero signal regions.

Spatial normalization: Resampling all volumes to 128×128×128 dimensions.

Intensity normalization: Z-score normalization of voxel intensities.

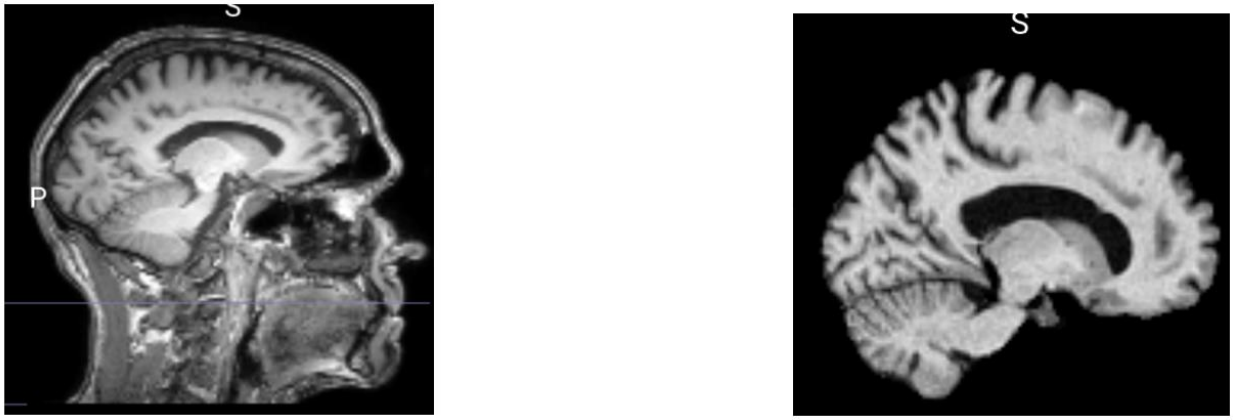

Figure 7.
Comparison before and after skull-stripping.

4. Experimental Results and Analysis

Single-modality binary (AD vs. CN) and three-class (AD vs. MCI vs. CN) experiments were evaluated across baseline and improved models, accompanied by ablation comparisons against published literature.

4.1 Evaluation Metrics

We used Accuracy (ACC), Sensitivity (SEN), and Specificity (SPE) to evaluate:

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

$$SEN = \frac{TP}{TP + FN} \times 100\% \quad (11)$$

$$SPE = \frac{TN}{TN + FP} \times 100\% \quad (12)$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively. ACC served as the primary overall metric.

4.2. Experimental Results

The experiment is divided into three parts: PET binary classification, PET ternary classification, and MRI classification to validate the model's classification performance across different medical imaging modalities and varying numbers of classes.

4.2.1. PET Binary Cl

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 8: 169-189, 2026
DOI: 10.55214/2576-8484.v10i8.13501

assification (AD vs. CN)

Results for PET binary classification are summarized in Table 3. Our improved CNN-Transformer model achieved the highest performance, with 92.54% ACC and 91.50% SEN.

Table 3.
PET Binary Classification Results.

Model	ACC (%)	SEN (%)	SPE (%)
CNN+SE ATTENTION	91.04	93.55	88.89
Transformer	89.55	87.10	91.67
CNN+Transformer (Unimproved)	91.04	90.32	91.67
CNN+Transformer (Improved)	92.54	91.50	90.79

First, the Transformer model without convolutional layers got 89.55% ACC, with 87.10% SEN and 91.67% SPE. The Transformer has a strong ability to learn global information and can capture relationships between different brain regions. However, PET images also contain many local metabolic features. Since the Transformer mainly focuses on global feature extraction, some local spatial information may not be fully learned. This may affect its classification performance. Therefore, its results are slightly lower than those of models combined with convolution operations.

The CNN+SE Attention model achieved 91.04% ACC. Its SEN and SPE were 93.55% and 88.89%, respectively. After adding the SE Attention module, the network could assign more attention to important feature channels. This helped the model focus on disease-related regions in PET images. As a result, this model obtained the highest sensitivity among the tested methods. However, CNN mainly extracts local image features. It has limitations when modeling relationships between distant brain regions. Therefore, the specificity of this model was relatively lower.

The CNN+Transformer model before improvement achieved an ACC of 91.04%, a SEN of 90.32%, and an SPE of 91.67%. Our integrated model performed better than the individual CNN model and the individual Transformer model. CNN extracts the spatial features of PET images, while the Transformer can capture the global dependencies between different regions. The combination of these two structures allows the model to obtain more complete image information. The specificity of 91.67% also indicates that the model has a good ability to distinguish normal samples.

The improved CNN+Transformer model proposed in this study achieved the best results. The ACC, SEN, and SPE were 92.54%, 91.50%, and 90.79%, respectively. The accuracy was increased by 1.50% over the original CNN+Transformer model. This improvement shows that the improved network structure can further improve the ability of feature learning.

In the improved model, a deeper residual convolutional structure was used for feature extraction. At the same time, the Transformer encoder was adjusted to improve feature interaction. Through these changes, local features of brain regions and global semantic features are combined more effectively. Therefore, the model obtained better representation ability for PET images.

The sensitivity of the improved model was slightly lower than that of the CNN+SE Attention model. However, it still remained above 90%, which means the model could effectively identify AD samples. Meanwhile, the specificity was 90.79%, which indicates a good recognition ability for the normal samples. The improved model had a better balance between the two indicators compared to models that focused only on sensitivity or specificity.

PET images reveal local metabolic alterations and connections between distinct brain areas. A single model structure cannot fully extract all useful information. CNNs are appropriate for learning local details, while Transformers are more effective at capturing global relationships. Fusing these two methods can enhance the performance of feature extraction and classification.

The enhanced CNN+Transformer model achieved higher classification accuracy and stable performance in PET image classification by optimizing the feature extraction process and fusion strategy. Results show its potential value for the auxiliary diagnosis of AD.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 8: 169-189, 2026
DOI: 10.55214/2576-8484.v10i8.13501

Overall, our improved CNN+Transformer model could classify PET brain images into AD and normal control groups. It provides a valuable methodology for the early diagnosis of Alzheimer's disease.

4.2.2. PET Three-Class Classification (AD vs. MCI vs. CN)

Three-class classification experiments were conducted on PET images to further examine the classification ability of different models under more complex conditions. The task included three categories: AD, MCI, and CN. ACC, SEN, and SPE were the evaluation indicators. These experimental results are shown in Table 4.

Table 4.
PET Three-Class Classification Experimental Results.

Model	ACC (%)	SEN (%)	SPE (%)
CNN+SE ATTENTION	46.82	65.55	62.10
Transformer	59.52	56.43	60.76
CNN+Transformer (Unimproved)	62.70	58.88	63.91
CNN+Transformer (Improved)	68.25	66.17	69.49

Compared with binary classification, the three-class task is more difficult because the differences among AD, MCI, and CN are relatively small. In particular, MCI represents an intermediate stage between CN and AD. The metabolic characteristics of MCI patients may overlap with those of both the CN and AD groups, which increases the difficulty of model classification. Therefore, the performance of the four models declined to different degrees in the three-class experiments.

The CNN+SE Attention model achieved an ACC of 46.82%, with SEN of 65.55% and SPE of 62.10%. The high relative sensitivity allows the model to detect some samples related to the disease. However, the overall accuracy was low. The model had difficulty discriminating among the three categories. One possible reason is that the SE Attention mechanism mainly enhances important feature channels through attention weighting. Although this improves the response to certain local regions, the model has limited capacity to identify connections between different brain regions. As a result, some metabolic changes related to the MCI stage may not be fully represented.

The Transformer model obtained an ACC of 59.52%, SEN of 56.43%, and SPE of 60.76%. The Transformer achieved better classification accuracy compared with the CNN+SE Attention model. This is mainly because the self-attention mechanism can establish connections between different image regions and capture global information. However, PET images contain many local metabolic features, and the Transformer does not have the same local feature extraction ability as convolutional networks. If training data is limited, this limitation may affect the quality of learned features.

The pre-improved CNN+Transformer model achieved an ACC of 62.70%, a SEN of 58.88%, and an SPE of 63.91%. After introducing convolutional layers, the model obtained stronger local feature extraction ability while maintaining the global modeling capability of the Transformer. The accuracy increased by 3.18% compared with the Transformer model. The rise in specificity also suggests that the model was more capable of separating CN samples from others.

The improved CNN+Transformer model performed best among all the methods tested. The accuracy was 68.25%, the sensitivity was 66.17%, and the specificity was 69.49%. ACC, SEN, and SPE increased by 5.55%, 7.29%, and 5.58%, respectively, compared with the pre-improved model. These results show that the improved structure further enhanced the model's capacity to extract features from PET.

In the improved network, a deeper 3D residual convolution structure was adopted to obtain more detailed local metabolic information. Meanwhile, the Transformer module was used to model relationships between different brain regions. The combination of these two components allowed it to obtain local and global features from PET images. This was especially helpful for MCI samples, whose imaging characteristics are often between those of AD and CN and are more difficult to distinguish.

Compared with the binary classification task, the performance improvement of our improved model was more obvious in three-class classification experiments. This suggests that the optimized

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 8: 169-189, 2026
DOI: 10.55214/2576-8484.v10i8.13501

CNN+Transformer structure has better adaptability when dealing with more complex classification problems.

CNNs are good at finding local details, but Transformers are better at seeing connections across the whole brain. Their combination gives a more complete feature representation. The improved CNN+Transformer model achieved better classification performance in AD, MCI, and CN discrimination. Its potential in deep learning-based auxiliary diagnosis of AD is great.

4.2.3. MRI Classification

To determine the effect of the improved model on the auxiliary diagnosis of AD, comparative experiments were conducted on binary and three-class tasks. Traditional convolutional networks, Transformers, the pre-improved model, and the improved model were evaluated, alongside horizontal comparisons with existing literature. The experimental results are presented in Table 5, and the confusion matrix is shown in Table 6.

Overall, the improved CNN+Transformer model constructed in this paper achieved peak performance across both binary and three-class tasks, demonstrating that the designed single-modality MRI deep learning framework effectively enhances Alzheimer's disease classification performance and exhibits particularly strong discriminative power in the more challenging three-class task.

Table 5.
MRI Experimental Results.

	AD vs. CN			AD vs. MCI vs. CN		
Model	**ACC(%)**	**SEN(%)**	**SPE(%)**	**ACC(%)**	**SEN(%)**	**SPE(%)**
CNN+SE ATTENTION	92.54	91.23	93.39	58.33	56.63	77.14
Transformer	88.05	89.81	91.04	55.83	53.59	75.65
CNN+Transformer(Unimproved)	92.54	91.23	93.39	61.67	62.03	79.66
Convolutional Block Attention [38]	80.23	-	-	59.35	-	-
Line-Hypergraph Neural Network [39] **Error! Reference source not found.**	81.50	-	-	64.20	-	-
Generalized Cross-Entropy Loss [40]	81.70	-	-	67.40	-	-
CNN+Transformer(improved)	94.03	92.57	94.80	69.16	66.73	81.27

Table 6.
Three-Class Confusion Matrix of Improved CNN+Transformer Model.

	CN	**MCI**	**AD**
CN	51.42%	45.71%	2.85%
MCI	7.54%	69.81%	22.64%
AD	0.00%	12.50%	87.50%

In the AD and CN binary classification experiment, the improved CNN+Transformer model got an ACC of 94.03%, a SEN of 92.57%, and an SPE of 94.80%. Among all comparison methods, this model obtained the best results in the three evaluation indicators. The CNN+SE Attention model had an accuracy of 92.54%. This showed that CNN combined with channel attention can effectively extract important local structural features from MRI. The Transformer model had a lower ACC of 88.05%. This may be related to the small number of medical image samples. Therefore, it is difficult for the Transformer to learn sufficient spatial feature representations when used independently.

The pre-improved CNN+Transformer model achieved an accuracy of 92.54%. Compared with the CNN or Transformer model, the combination of the two structures improved feature-learning abilities by integrating local information and global dependencies. The improved CNN+Transformer model obtained an ACC of 94.03%, which was 1.49% higher than the pre-improved model.

The experimental results indicate that CNN and Transformer are complementary in MRI image analysis. CNN can learn local anatomical structures, and Transformer can model the relationships among different brain regions. Mixing these two parts helps the model obtain more information from the images.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 8: 169-189, 2026
DOI: 10.55214/2576-8484.v10i8.13501

Our model achieved a sensitivity of 92.57% and a specificity of 94.80%. It can distinguish AD from CN people without favoring one side.

Sorting images into three groups (AD, MCI, and CN) is harder than sorting them into just two because MCI is an intermediate stage. Thus, there are few differences between CN and MCI, which makes them hard for the model to distinguish. In that task, the CNN+SE model achieved 58.33% ACC. The Transformer model achieved 55.83% ACC. Our first mix of CNN+Transformer got 61.67% ACC. The improved model got 69.16% ACC. Also, it had a sensitivity of 66.73% and specificity of 81.27%. This shows it still performs well in this task.

Compared with traditional CNN models, the proposed method not only extracts local image features but also considers the relationships between different brain regions through a Transformer. This helps the model capture subtle differences among MCI, AD, and CN samples.

For medical image classification tasks, models based only on Transformers may face difficulties when the training data size is limited. Meanwhile, traditional CNN models are restricted by their local receptive fields and have limited capacity to capture a wide range of information. The CNN+Transformer structure has the advantages of both methods and improves the capacity of feature representation in multiclass classification scenarios.

To improve the effect of the proposed model, additional comparisons were conducted with existing studies. In the three-class classification task, Convolutional Block Attention achieved an accuracy of 59.35%, Line-Hypergraph Neural Network achieved 64.20%, and Generalized Cross-Entropy Loss achieved 67.40%. Compared with them, the improved CNN+Transformer model achieved the highest accuracy of 69.16%. It was 1.76% higher than Generalized Cross-Entropy Loss and 9.81% higher than Convolutional Block Attention.

The results show our model works well with just MRI scans. Some other studies use very complex steps, but we kept our network simple by mixing CNNs and Transformers.

4.3. Analysis of Experimental Results

Results show that our enhanced CNN+Transformer model is effective for all tasks. It works well at picking out important features from the MRI and is better than the.

The model achieved 94.03% accuracy on the two-group task. Sensitivity was 92.57%, and specificity was 94.80%. The numbers went up by about 1.4%-1.5% from the unimproved model. This means the new model is better at telling AD patients apart from healthy people.

On the harder three-group task (AD, MCI, and CN), the model achieved 69.16% accuracy. Sensitivity was 66.73%, and specificity was 81.27%. Accuracy rose by 7.49%. This shows our model is good at sorting similar groups.

We also looked at how the parts work together. By using both CNN and transformers, the model gets a fuller picture of the image. So, it is more stable when things get complicated.

Overall, the improved model works well in classifying AD and MCI. The results prove that this method is useful for helping diagnose AD using MRI. It also gives other researchers a good example to follow.

5. Conclusion and Future Directions

Existing models had trouble seeing both small details and the big picture. Therefore, we built a hybrid model that combines CNNs and Transformers. It can capture local brain features and global connections at the same time and shows good results.

5.1. Conclusions

In this study, a three-dimensional CNN-Transformer hybrid network was developed for AD classification using structural MRI and PET. The main research work and conclusions are summarized as follows.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 8: 169-189, 2026
DOI: 10.55214/2576-8484.v10i8.13501

(1) A standardized MRI image preprocessing procedure was established. The original MRI data were processed through several steps, including image format conversion, spatial normalization, and intensity adjustment. These procedures eliminated differences due to data acquisition conditions and provided a uniform input format for subsequent model training.

(2) A hybrid CNN-Transformer network was proposed by combining convolutional feature extraction and global modeling using a transformer. We adopted the CNN module to obtain spatial brain-structure features and proposed the Transformer module to learn the relationships among different brain regions. This structure improved the representation ability of MRI image features.

(3) The network structure was further optimized based on the characteristics of medical image classification. Improvements mainly focused on residual feature extraction and Transformer feature modeling, enabling the model to obtain complete image representations and improving classification performance.

(4) Comparative experiments were used to evaluate the effectiveness of the proposed model. The improved CNN-Transformer model achieved an ACC of 94.03% in binary classification and 69.16% in the three-class task. This indicates that the improved model can distinguish different categories in Alzheimer's disease-related MRI.

5.2. Key Contributions

This study focuses on auxiliary AD diagnosis using deep learning and explores the application of CNN–Transformer hybrid structures in MRI classification. Its main contributions are summarized as follows.

(1) Analysis of recent research methods in AD auxiliary diagnosis.

This study reviewed the pathological characteristics of AD, MRI imaging methods, and diagnostic approaches using deep learning. Existing CNN, Transformer, and medical image classification methods were reviewed, particularly their advantages and disadvantages in feature extraction and information modeling. This review provides a theoretical basis for the model design in this study.

(2) Construction of a CNN-Transformer hybrid classification model for MRI images.

Considering that CNNs have advantages in extracting local spatial features but have limited ability to model long-range relationships, this study combines CNNs with Transformers. The CNN module extracts detailed structural information from MRI images, while the Transformer captures global dependencies among different regions. Through the combination of these two modules, the model can obtain richer feature representations.

(3) Verification of the effectiveness of the hybrid structure in AD classification.

Based on MRI image datasets, classification experiments were conducted using different network structures. The results reveal that our mixed model outperformed models using only a CNN or only a Transformer. This indicates that leveraging local and global information is a promising approach for classifying brain images.

(4) Exploration of deep learning model design for medical image analysis.

Experiments show that the system learns better by comparing different models and mixing them. This gives other researchers a good idea of how to build medical imaging tools.

(5) Providing experimental support for intelligent auxiliary diagnosis of Alzheimer's disease.

Our model learns from MRI scans by itself. It sorts people into AD, MCI, or normal groups. This study shows that deep learning can help diagnose Alzheimer's. It also gives other researchers a good starting point for future work.

5.3. Limitations and Future Recommendations

Our model performed well in the tests, but it is not perfect. There are still some issues in the data and in our model-building. "We have to do more work before doctors can use it in real life." In what follows, we discuss these problems and our plans for the next steps.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 8: 169-189, 2026
DOI: 10.55214/2576-8484.v10i8.13501

5.3.1. Limitations

(1) Limited dataset size

Due to the difficulty of obtaining medical imaging data, the dataset used in this study was relatively limited. A small sample size may affect the learning ability of deep learning models and reduce their robustness when dealing with data from different populations.

(2) Lack of multimodal information integration

This study mainly focused on structural MRI images. Although MRI provides important information about brain structure, other types of information, such as PET images, functional MRI, and clinical assessment indicators, may provide additional disease-related information. The absence of multimodal data limits further improvement of diagnostic performance.

(3) Limited model interpretability

Although Transformer attention mechanisms can provide certain information about important image regions, the internal decision-making process of deep learning models remains difficult to explain. More effective visualization and interpretation methods are needed to improve clinical acceptance.

(4) Insufficient external validation

The experiments were mainly conducted on public datasets. The model's generalization capacity across different medical centers, imaging devices, and scanning protocols requires further validation.

5.3.2. Recommendations for Future Research

(1) Multimodal data fusion

Future studies can combine structural MRI with PET, functional MRI, and clinical information. By using multimodal fusion methods, complementary information from different sources is utilized to improve early diagnostic performance.

(2) Improving model interpretability

Explainable artificial intelligence methods, such as Grad-CAM and Integrated Gradients, can be introduced to visualize important brain regions contributing to classification decisions.

(3) Multi-center clinical validation

In the future, we will have to gather data from many different hospitals. This can tell us how the model performs in the real world. The model will become more reliable and ready for real-world use.

Transparency:

The author confirms that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Copyright:

References

[1] P. Q. Gong, Z. J. Yan, and X. Li, "Image classification method of Alzheimer's disease based on deep learning," *Chinese Journal of Medical Physics*, vol. 42, no. 11, pp. 1420-1433, 2025.

[2] D. Lu, K. Popuri, G. W. Ding, R. Balachandar, and M. F. Beg, "Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images," *Scientific Reports*, vol. 8, p. 5697, 2018. https://doi.org/10.1038/s41598-018-22871-z

[3] W. Lin *et al.*, "Bidirectional mapping of brain MRI and PET with 3D reversible GAN for the diagnosis of Alzheimer's disease," *Frontiers in Neuroscience*, vol. 15, p. 646013, 2021. https://doi.org/10.3389/fnins.2021.646013

[4] J. Song, J. Zheng, P. Li, X. Lu, G. Zhu, and P. Shen, "An effective multimodal image fusion method using MRI and PET for Alzheimer's disease diagnosis," *Frontiers in Digital Health*, vol. 3, p. 637386, 2021. https://doi.org/10.3389/fdgth.2021.637386

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 8: 169-189, 2026
DOI: 10.55214/2576-8484.v10i8.13501

[5] M. Liu, D. Cheng, K. Wang, Y. Wang, and A. S. D. N. Initiative, "Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis," *Neuroinformatics*, vol. 16, no. 3, pp. 295-308, 2018. https://doi.org/10.1007/s12021-018-9370-4

[6] C. Feng *et al.*, "Deep learning framework for Alzheimer's disease diagnosis via 3D-CNN and FSBi-LSTM," *IEEe Access*, vol. 7, pp. 63605-63618, 2019. https://doi.org/10.1109/ACCESS.2019.2913847

[7] Y. Huang, J. Xu, Y. Zhou, T. Tong, X. Zhuang, and A. s. D. N. Initiative, "Diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network," *Frontiers in Neuroscience*, vol. 13, p. 509, 2019. https://doi.org/10.3389/fnins.2019.00509

[8] M. Narazani, I. Sarasua, S. Pölsterl, A. Lizarraga, I. Yakushev, and C. Wachinger, "Is a pet all you need? A multi-modal study for alzheimer's disease using 3d cnns," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 66-76). Cham: Springer Nature Switzerland*, 2022.

[9] W. Li, Y. Zhang, G. Wang, Y. Huang, and R. Li, "DFENet: A dual-branch feature enhanced network integrating transformers and convolutional feature learning for multimodal medical image fusion," *Biomedical Signal Processing and Control*, vol. 80, p. 104402, 2023. https://doi.org/10.1016/j.bspc.2022.104402

[10] Y. Zhang, X. He, Y. Liu, C. Z. L. Ong, Y. Liu, and Q. Teng, "An end-to-end multimodal 3D CNN framework with multi-level features for the prediction of mild cognitive impairment," *Knowledge-Based Systems*, vol. 281, p. 111064, 2023. https://doi.org/10.1016/j.knosys.2023.111064

[11] X. Gao, F. Shi, D. Shen, and M. Liu, "Multimodal transformer network for incomplete image generation and diagnosis of Alzheimer's disease," *Computerized Medical Imaging and Graphics*, vol. 110, p. 102303, 2023. https://doi.org/10.1016/j.compmedimag.2023.102303

[12] S. Miao *et al.*, "MMTFN: Multi-modal multi-scale transformer fusion network for Alzheimer's disease diagnosis," *International Journal of Imaging Systems and Technology*, vol. 34, no. 1, p. e22970, 2024. https://doi.org/10.1002/ima.22970

[13] Y. Tang, X. Xiong, G. Tong, Y. Yang, and H. Zhang, "Multimodal diagnosis model of Alzheimer's disease based on improved Transformer," *Biomedical Engineering Online*, vol. 23, no. 1, p. 8, 2024. https://doi.org/10.1186/s12938-024-01204-4

[14] M. Kavitha, N. Yudistira, and T. Kurita, "Multi instance learning via deep CNN for multi-class recognition of Alzheimer's disease," in *2019 IEEE 11th International Workshop on Computational Intelligence and Applications (IWCIA) (pp. 89-94). IEEE*, 2019.

[15] X. Bi, S. Li, B. Xiao, Y. Li, G. Wang, and X. Ma, "Computer aided Alzheimer's disease diagnosis by an unsupervised deep learning technology," *Neurocomputing*, vol. 392, pp. 296-304, 2020. https://doi.org/10.1016/j.neucom.2018.11.111

[16] H. Jia, Y. Wang, Y. Duan, and H. Xiao, "Alzheimer's disease classification based on image transformation and features fusion," *Computational and Mathematical Methods in Medicine*, vol. 2021, no. 1, p. 9624269, 2021. https://doi.org/10.1155/2021/9624269

[17] R. Sharma, T. Goel, M. Tanveer, and R. Murugan, "FDN-ADNet: Fuzzy LS-TWSVM based deep learning network for prognosis of the Alzheimer's disease using the sagittal plane of MRI scans," *Applied Soft Computing*, vol. 115, p. 108099, 2022. https://doi.org/10.1016/j.asoc.2021.108099

[18] A. Rogeau *et al.*, "A 3D convolutional neural network to classify subjects as Alzheimer's disease, frontotemporal dementia or healthy controls using brain 18F-FDG PET," *Neuroimage*, vol. 288, p. 120530, 2024. https://doi.org/10.1016/j.neuroimage.2024.120530

[19] Z. Qiu *et al.*, "3D multimodal fusion network with disease-induced joint learning for early Alzheimer's disease diagnosis," *IEEE Transactions on Medical Imaging*, vol. 43, no. 9, pp. 3161-3175, 2024. https://doi.org/10.1109/TMI.2024.3386937

[20] B. Cheng, M. Liu, D. Shen, Z. Li, D. Zhang, and A. S. D. N. Initiative., "Multi-domain transfer learning for early diagnosis of Alzheimer's disease," *Neuroinformatics*, vol. 15, no. 2, pp. 115-132, 2017. https://doi.org/10.1007/s12021-016-9318-5

[21] S. S. Kumar and M. Nandhini, "Entropy slicing extraction and transfer learning classification for early diagnosis of Alzheimer diseases with sMRI," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 17, no. 2, pp. 1-22, 2021. https://doi.org/10.1145/3383749

[22] S. Naz, A. Ashraf, and A. Zaib, "Transfer learning using freeze features for Alzheimer neurological disorder detection using ADNI dataset: S. Naz et al," *Multimedia Systems*, vol. 28, no. 1, pp. 85-94, 2022. https://doi.org/10.1007/s00530-021-00797-3

[23] A. Mehmood *et al.*, "A transfer learning approach for early diagnosis of Alzheimer's disease on MRI images," *Neuroscience*, vol. 460, pp. 43-52, 2021. https://doi.org/10.1016/j.neuroscience.2021.01.002

[24] M. Odusami, R. Maskeliūnas, R. Damaševičius, and T. Krilavičius, "Analysis of features of Alzheimer's disease: Detection of early stage from functional brain changes in magnetic resonance images using a finetuned ResNet18 network," *Diagnostics*, vol. 11, no. 6, p. 1071, 2021. https://doi.org/10.3390/diagnostics11061071

[25] P. Buvaneswari and R. Gayathri, "Deep learning-based segmentation in classification of Alzheimer's disease," *Arabian Journal for Science and Engineering*, vol. 46, no. 6, pp. 5373-5383, 2021. https://doi.org/10.1007/s13369-020-05193-z

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 8: 169-189, 2026
DOI: 10.55214/2576-8484.v10i8.13501

[26] M. Odusami, R. Maskeliūnas, and R. Damaševičius, "An intelligent system for early recognition of Alzheimer's disease using neuroimaging," *Sensors*, vol. 22, no. 3, p. 740, 2022. https://doi.org/10.3390/s22030740

[27] A. Ashraf, S. Naz, S. H. Shirazi, I. Razzak, and M. Parsad, "Deep transfer learning for alzheimer neurological disorder detection," *Multimedia Tools and Applications*, vol. 80, no. 20, pp. 30117-30142, 2021. https://doi.org/10.1007/s11042-020-10331-8

[28] M. W. Oktavian, N. Yudistira, and A. Ridok, "Classification of Alzheimer's disease using the Convolutional Neural Network (CNN) with transfer learning and weighted loss," *arXiv preprint arXiv:2207.01584*, 2022. https://doi.org/10.48550/arXiv.2207.01584

[29] N. Gharaibeh, A. Abu-Ein, O. M. Al-hazaimeh, K. M. Nahar, W. A. K. Abu-Ain, and M. Al-Nawashi, "Swin transformer-based segmentation and multi-scale feature pyramid fusion module for Alzheimer's disease with machine learning," *International Journal of Online and Biomedical Engineering*, vol. 19, no. 4, pp. 22-50, 2023.

[30] M. Leela, K. Helenprabha, and L. Sharmila, "Prediction and classification of Alzheimer disease categories using integrated deep transfer learning approach," *Measurement: Sensors*, vol. 27, p. 100749, 2023. https://doi.org/10.1016/j.measen.2023.100749

[31] M. A. Mahmud Joy, S. Nasrin, A. Siddiqua, and D. M. Farid, "ViTAD: Leveraging modified vision transformer for Alzheimer's disease multi-stage classification from brain MRI scans," *Brain Research*, vol. 1847, p. 149302, 2025. https://doi.org/10.1016/j.brainres.2024.149302

[32] R. Turrisi, S. Pati, G. Pioggia, G. Tartarisco, and A. S. D. N. Initiative, "Adapting to evolving MRI data: A transfer learning approach for Alzheimer's disease prediction," *NeuroImage*, vol. 307, p. 121016, 2025. https://doi.org/10.1016/j.neuroimage.2025.121016

[33] H. Li *et al.*, "A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data," *Alzheimer's & Dementia*, vol. 15, no. 8, pp. 1059-1070, 2019. https://doi.org/10.1016/j.jalz.2019.02.007

[34] E. Lee, J.-S. Choi, M. Kim, and H.-I. Suk, "Toward an interpretable Alzheimer's disease diagnostic model with regional abnormality representation via deep learning," *NeuroImage*, vol. 202, p. 116113, 2019. https://doi.org/10.1016/j.neuroimage.2019.116113

[35] J. Li, L. Liu, and H. Wu, "Multi-scale channel-spatial perception Mamba for Alzheimer's disease PET image classification method," *Journal of Guangdong University of Technology*, vol. 43, no. 1, pp. 31–39, 2026.

[36] Z. W. Xiao, J. L. Bi, and H. Y. He, "ResNet50 model integrating DW convolution and SE attention for predicting Alzheimer's disease from MRI," *Computer Era*, vol. 2, pp. 28–34, 2026.

[37] W. Lin, "Synthesizing missing data using 3D reversible GAN for Alzheimer's disease," in *Proceedings of the 2020 International Symposium on Artificial Intelligence in Medical Sciences (ISAIMS 2020) (pp. 208–213). Association for Computing Machinery*, 2020.

[38] S. Liu, C. Yadav, C. Fernandez-Granda, and N. Razavian, "On the design of convolutional neural networks for automatic detection of Alzheimer's disease," in *Machine Learning for Health Workshop (pp. 184-201). PMLR*, 2020.

[39] Y. J. Su, Y. Li, and P. J. Cao, "Alzheimer's disease classification based on line-hypergraph neural network," *Computer Systems & Applications*, vol. 32, no. 6, pp. 260-268, 2023.

[40] J. Wen *et al.*, "Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation," *Medical Image Analysis*, vol. 63, p. 101694, 2020. https://doi.org/10.1016/j.media.2020.101694

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 8: 169-189, 2026
DOI: 10.55214/2576-8484.v10i8.13501