

RESEARCH

Open Access



MHAGuideNet: a 3D pre-trained guidance model for Alzheimer's Disease diagnosis using 2D multi-planar sMRI images

Yuanbi Nie¹, Qiushi Cui^{1*}, Wenyuan Li¹, Yang Lü² and Tianqing Deng³

Abstract

Background Alzheimer's Disease is a neurodegenerative condition leading to irreversible and progressive brain damage, with possible features such as structural atrophy. Effective precision diagnosis is crucial for slowing disease progression and reducing the incidence rate and morbidity. Traditional computer-aided diagnostic methods using structural MRI data often focus on capturing such features but face challenges, like overfitting with 3D image analysis and insufficient feature capture with 2D slices, potentially missing multi-planar information, and the complementary nature of features across different orientations.

Methods The study introduces MHAGuideNet, a classification method incorporating a guidance network utilizing multi-head attention. The model utilizes a pre-trained 3D convolutional neural network to direct the feature extraction of multi-planar 2D slices, specifically targeting the detection of features like structural atrophy. Additionally, a hybrid 2D slice-level network combining 2D CNN and 2D Swin Transformer is employed to capture the interrelations between the atrophy in different brain structures associated with Alzheimer's Disease.

Results The proposed MHAGuideNet is tested using two datasets: the ADNI and OASIS datasets. The model achieves an accuracy of 97.58%, specificity of 99.89%, F1 score of 93.98%, and AUC of 99.31% on the ADNI test dataset, demonstrating superior performance in distinguishing between Alzheimer's Disease and cognitively normal subjects. Furthermore, testing on the independent OASIA test dataset yields an accuracy of 96.02%, demonstrating the model's robust performance across different datasets.

Conclusion MHAGuideNet shows great promise as an effective tool for the computer-aided diagnosis of Alzheimer's Disease. Within the guidance of information from the 3D pre-trained CNN, the ability to leverage multi-planar information and capture subtle brain changes, including the interrelations between different structural atrophies, underscores its potential for clinical application.

Keywords Alzheimer's Disease, Structural MRI, Computer-aided diagnosis, Image classification

*Correspondence:

Qiushi Cui
qcui@cqu.edu.cn

¹ School of Electrical Engineering, Chongqing University, Shapingba, Chongqing 400044, China

² Department of Geriatrics, The First Affiliated Hospital of Chongqing Medical University, Youyi Road, Chongqing 401122, China

³ Department of Geriatrics, Chongqing Medical University, Medical College Road, Chongqing 400016, China

Introduction

Alzheimer's Disease (AD) is one of the most prevalent progressive neurological diseases and a primary cause of dementia [1]. AD typically causes prominent amnesic cognitive impairment. The presentation with short-term memory difficulty is the most common symptom but the impairments in expressive speech, visuospatial processing, and executive functions also often occur [2]. Since no cure has been developed for AD yet, the current



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

treatment is focused on reducing the progress speed to a severe stage. Therefore, an early precision diagnosis is crucial for slowing the disease’s development and reducing the incidence rate and morbidity.

Clinically, sMRI provides high-resolution structural images of the brain, enabling the observation of anatomical and functional neural changes associated with AD. In clinical practice, experienced radiologists visually assess the sMRI images using standardized rating scales to determine the extent of atrophy in specific areas such as the medial temporal lobe, posterior brain, or entire brain. For example, the Medial Temporal Atrophy (MTA) score, illustrated in Fig. 1, reflects the severity of atrophy in the medial temporal lobe [3]. However, the manual analysis of the images is not only time-consuming but also inherently subjective, heavily relying on the abilities and experience of a clinician. Additionally, processing massive and complex sMRI images is a challenging task for clinicians.

With the development of machine learning (ML) techniques, many algorithms employing medical imaging have emerged for computer-aided diagnosis (CAD) applications. For instance, DermoNet [4] has shown significant improvements in dermoscopic disease recognition, highlighting the effectiveness of deep learning models in medical image analysis. Building on these advances, numerous CAD algorithms for AD utilizing sMRI have also been proposed. Traditional ML approaches split brain sMRI images into different regions of interest (ROIs) using a structural template and apply specific algorithms to discern structural variations or volume changes in the areas crucial to AD progression [5–7]. Subsequently, these features are input into an ML model for classification, such as a random forest or support vector machine. However, this approach requires a manual feature selection procedure, limiting the model’s ability to capture complex patterns in the data. In contrast, convolutional neural networks (CNNs), which is a deep learning model, autonomously derive discriminative features from sMRI images, providing superior generalization capabilities [8–11]. This

automatic feature extraction is advantageous for capturing intricate patterns in medical images, which is essential for accurate disease diagnosis. Despite their efficacy, however, CNNs face limitations in learning contextual information. This results in difficulties in recognizing long-range dependencies and correlations among distant anatomical regions in sMRI analysis. Such challenges arise because critical regional changes are not isolated but intricately interconnected. To overcome the limitations of CNNs, some research has focused on integrating attention mechanisms to enhance feature extraction. Bakkouri et al. [12] proposed a multi-scale feature extraction and attention-based network, which demonstrates the effectiveness of combining multi-level representations with attentional mechanisms for downstream tasks. Furthermore, novel Transformer-based models such as the Vision Transformer (ViT) [13] and Swin Transformer [14] have been introduced, yielding impressive results compared to CNNs. The Swin Transformer demonstrates an improved performance over traditional CNNs by more effectively integrating local details, long-range dependencies, and global contextual information, which is vital in comprehending the complex structure of brain images. In addition, some literature has proposed models that combine CNN and Transformer. In [15], a 2.5D-subject method and a two-stream structure combining CNN and Swin Transformer were developed to enhance the ability to capture features.

Recent advancements have led to the development of 3D CNN, which facilitates feature extraction at the subject-level across the entire image, as opposed to the slice-level analysis characteristic of 2D CNNs [16]. While this approach is promising for capturing the spatial relationships in 3D data, it also faces challenges such as a higher propensity for overfitting on smaller datasets and longer training durations due to increased computational demands. Notably, each slice of the sMRI image across different planes contains significant local information, as shown in Fig. 2. Therefore, considering both global information at the

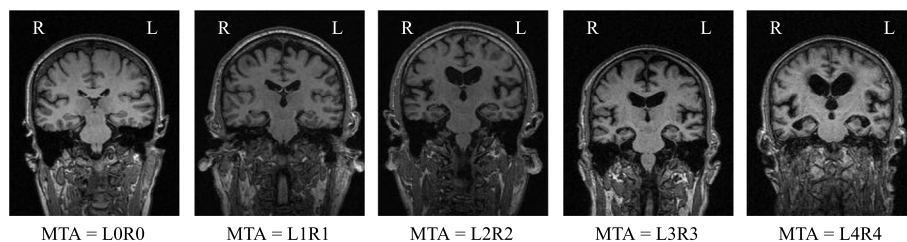


Fig. 1 Visual example of medial temporal atrophy(MTA) in different stages of brain atrophy where the L denotes left and the R denotes right. The MTA score is graded ranging from 0 to 4, with a higher score indicating a greater degree of atrophy

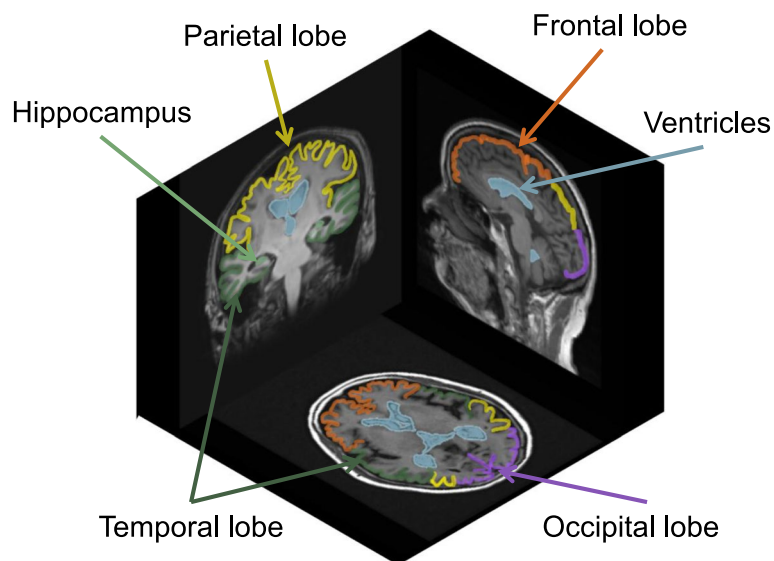


Fig. 2 The information of the regions of interest in different planes, including sagittal plane, coronal plane and axial plane

subject-level and diverse local information at the slice-level could provide valuable insights. Global information is instrumental in comprehending changes in the overall brain structures, whereas local information is crucial for detecting subtle alterations in specific areas. Considering both is essential for the accurate diagnosis of AD.

In this study, to address the issue of excessive reliance on doctors' experience and skills in the clinical diagnosis of AD, we design a CAD method called MHAGuideNet (Multi-head Attention Guide Network). This method overcomes some of the limitations found in previous CAD methods. CNN-based methods have limitations in learning feature dependencies, while the lesion areas in patients show certain correlations. Transformer-based methods are less capable of capturing subtle local changes in AD lesion areas compared to CNNs. Additionally, previous studies have focused on extracting features from the entire image or specific directional slices, ignoring the complementarity between features from different perspectives. Moreover, due to the relatively small dataset of medical images, it is difficult to capture lesions through equal treatment of all input features in limited iterative training. The MHAGuideNet harnesses the strengths of the pre-trained 3D CNN model to guide the feature extraction process in 2D slices from coronal, sagittal, and axial planes. This strategy not only augments the model's ability to comprehend intricate brain structures but also mitigates computational demands to a significant degree.

The main contributions of this study are summarized as follows:

(1) To enable rapid and accurate CAD of Alzheimer's Disease, we propose MHAGuideNet, which integrates a guidance network. This guidance component skillfully utilizes features extracted from 3D images to direct the 2D slices feature extraction process. The significance of this guidance strategy lies in its ability to detect detailed information in multi-planar 2D slices while concurrently encapsulating the complex spatial relationships inherent in 3D images. This approach affords a more holistic and nuanced understanding of the atrophy and abnormalities associated with AD.

(2) For a more nuanced and precise analysis of sMRI images, we deploy a multi-head attention mechanism within the guidance network. This mechanism selectively focuses on the most salient regions within the 3D feature map, effectively directing the analysis of subsequent 2D slices toward the regions of interest. Such precision in guiding the feature extraction process is pivotal for synthesizing a comprehensive and accurate brain representation.

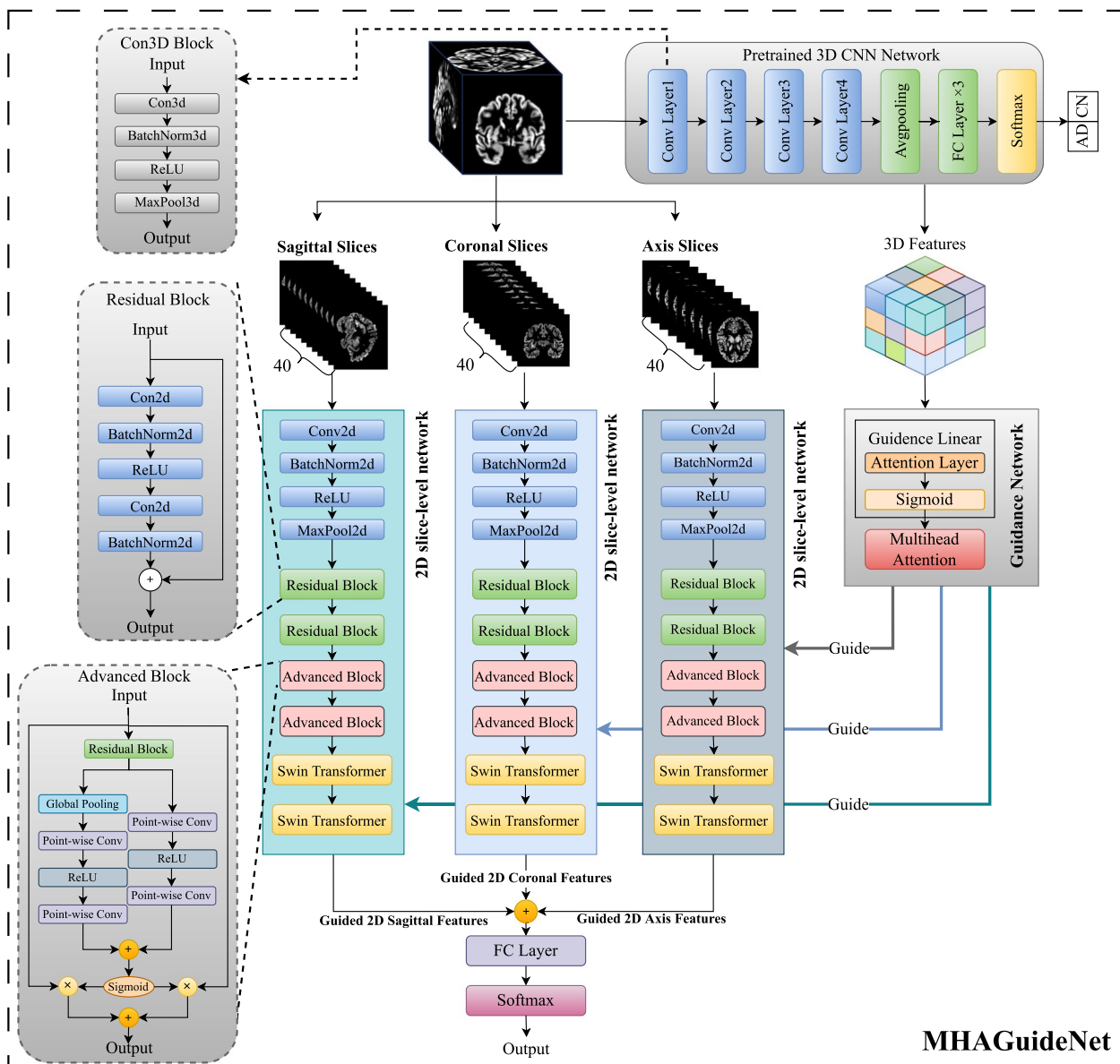
(3) Integrating local details and long-range dependencies is essential for capturing the complex correlations of brain structure within the 2D slice. To achieve this, we combine the 2D CNN and the 2D Swin Transformer with the attentional feature fusion mechanism in the 2D slice-level network of MHAGuideNet. This approach harnesses the CNNs' proficiency in extracting local and fine-grained features at the slice-level, while the Swin Transformer excels in capturing long-range dependencies and contextual information. The

synergy of these technologies in our network facilitates a more in-depth, contextually enriched analysis of sMRI images, leading to a potentially more accurate and reliable diagnosis of Alzheimer’s Disease.

Method

Our proposed model is a hybrid deep learning system combining a guidance mechanism with 2D slice-level feature processing, informed by 3D image features from a pre-trained 3D CNN. The model architecture, depicted in Fig. 3, a pre-trained 3D CNN guides the

2D network to focus on significant regions and the 2D slice-level network integrates 2D CNN and 2D Swin Transformer modules to extract the planar features of the slices and establish semantic connections using contextual information for capturing relationship features across different regions of an image. The final output involves concatenating the guided 2D features in a fully connected layer, followed by a softmax layer for categorization probabilities. In the Algorithm 1, we provide a detailed example to illustrate the application of our method.



MHAGuideNet

Fig. 3 The overall architecture of the proposed MHAGuideNet, including pre-trained 3D CNN network, 2D slice-level network and novel guidance network

Algorithm 1 MHAGuideNet for AD diagnosis

Input: \mathcal{X} : T1-weighted sMRI images
Output: \hat{y} : Classification result (AD or CN)

- 1: **Pre-trained 3D CNN Network**
- 2: **for** $b \in \{1, 2, 3, 4\}$ **do**
- 3: $\mathbf{H}_b \leftarrow \text{Conv3D}(\mathbf{H}_{b-1})$
- 4: $\mathbf{H}_b \leftarrow \text{BN3D}(\mathbf{H}_b)$
- 5: $\mathbf{H}_b \leftarrow \text{ReLU}(\mathbf{H}_b)$
- 6: $\mathbf{H}_b \leftarrow \text{MaxPool3D}(\mathbf{H}_b)$
- 7: **end for**
- 8: 3D feature maps: $\mathbf{z} \leftarrow \text{AvgPool3D}(\mathbf{H}_4)$
- 9: **Guidance Network Application**
- 10: $\mathbf{a} \leftarrow \text{AttentionLayer}(\mathbf{z})$
- 11: $\mathbf{a} \leftarrow \sigma(\mathbf{a})$
- 12: $\mathbf{a} \leftarrow \text{MultiheadAttention}(\mathbf{a})$
- 13: **2D Slice-Level Feature Extraction**
- 14: **for** $i \in \mathcal{X}$ **do**
- 15: $\mathcal{S}_{\text{sag}} \leftarrow \text{MiddleSlices}_{\text{Auto}}(i, \text{sagittal}, 40)$
- 16: $\mathcal{S}_{\text{cor}} \leftarrow \text{MiddleSlices}_{\text{Auto}}(i, \text{coronal}, 40)$
- 17: $\mathcal{S}_{\text{axi}} \leftarrow \text{MiddleSlices}_{\text{Auto}}(i, \text{axial}, 40)$
- 18: **for** $s \in \mathcal{S}_{\text{sag}} \cup \mathcal{S}_{\text{cor}} \cup \mathcal{S}_{\text{axi}}$ **do**
- 19: $\mathbf{f}_s \leftarrow \text{CNN_SwinTransformer}(s, \mathbf{a})$
- 20: **end for**
- 21: **end for**
- 22: **Feature Fusion and Classification**
- 23: $\mathbf{F} \leftarrow [\mathbf{f}_{s_1}, \mathbf{f}_{s_2}, \dots, \mathbf{f}_{s_{120}}]$
- 24: $\mathbf{h} \leftarrow \text{FC}(\mathbf{F})$
- 25: $\hat{y} \leftarrow \text{Softmax}(\mathbf{h})$

Pre-trained 3D CNN network

To extract information including volumetric information and intricate spatial relationships from 3D image data to guide the 2D slice-level feature extracting, we use a pre-trained 3D CNN network. This technique effectively captures spatial correlations across all three dimensions,

making it ideal for analyzing volumetric sMRI data. In the 3D CNN architecture, the design incorporates a series of distinct blocks. Each of these blocks is composed of several layers: a 3D convolution layer, followed by a 3D Batch Normalization (BN) layer, a ReLU activation layer, and culminating with a 3D max pooling layer. This sequential arrangement is repeated across all four blocks. After these blocks, a 3D average pooling layer condenses the multi-channel feature maps into a singular vector to encapsulate the global information derived from the preceding layers. We pre-train the network to classify AD and CN and subsequently utilize the output of average pooling as the 3D feature for guidance.

Guidance network

The 3D image data contains rich spatial information that is crucial for diagnosing Alzheimer’s Disease. However, using 3D networks also presents certain risks. While the 3D CNN network excels at capturing the spatial relationships in 3D data that are missed by 2D slices, it faces challenges such as a higher propensity for overfitting on smaller datasets and longer training times due to increased computational demands. To address these issues, we propose a guidance network that leverages the 3D information captured by the pre-trained 3D CNN to enhance 2D slice-level feature extraction. An example of this process is illustrated in Fig. 4. The visualization demonstrates how 3D features extracted by the 3D CNN are processed by the guidance network to generate attention features. These attention features subsequently guide the extraction of 2D features. The heatmap reveals that the regions of interest in the 3D features correspond to the highlighted areas in the 2D feature maps.

As shown in Fig. 3, the guidance network comprises two main components: the guidance linear block and the multi-head attention mechanism. In the guidance linear block, the output of the 3D CNN network $X_{in} = [x_1, x_2, \dots, x_M]^T \in \mathbb{R}^{M \times C}$ is transformed into an attention vector which is represented as $\Phi = [\phi_1, \phi_2, \dots, \phi_N]^T$, where N is the output dimension of the linear layer. Here,

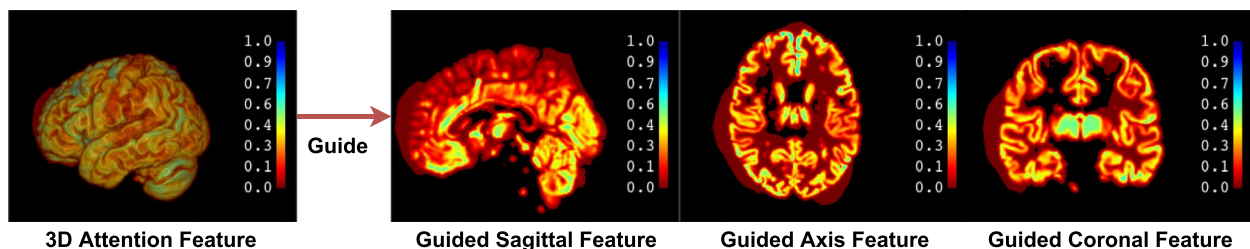


Fig. 4 Heatmap showing guidance of 2D feature extraction by 3D features. The visualization illustrates the process where 3D features extracted by 3D CNN are processed by a guidance network to obtain attention features. These attention features then guide the extraction of 2D features. The heatmap shows that the regions of interest in the 3D features correspond to the highlighted areas in the 2D feature maps

$x_m \in \mathbb{R}^{1 \times C}$, with M being the number of features and C the dimension of each feature in the 3D feature space. The transformation by the linear attention layer is crucial as it reduces the dimensionality of the 3D features from $M \times C$ to N , making them more manageable and suitable for guiding the 2D feature extraction process. The mathematical formulation of this layer is as follows:

$$\Phi = X_{in}W_a + b \quad W_a \in \mathbb{R}^{C \times N}; b \in \mathbb{R}^N, \tag{1}$$

where W_a is the weight of and b is the bias vector. To convert the attention vector into a guidance signal Ψ , a softmax activation function is applied as follows:

$$\Psi = \text{Softmax}(\Phi) \quad \Psi \in \mathbb{R}^N. \tag{2}$$

The softmax function normalizes the attention vector. The resulting guidance signal emphasizes the most

significant features of the 3D images. To further refine the guidance process, a multi-head attention mechanism is employed as shown in Fig. 5. This layer facilitates a complex, nuanced interaction between the 3D and 2D feature spaces. The attention mechanism dynamically adjusts to the input data, allowing the model to focus on the most relevant spatial features extracted from the 3D data for more precisely directing the processing of subsequent 2D slices towards the regions of importance. When employing a h-head multi-head attention mechanism, the input guidance signal Ψ is first mapped into queries (Q), keys (K), and values (V), with each mapping defined by the corresponding weight matrices (W_Q, W_K, W_V). This process can be expressed through the following formulas:

$$Q = \Psi \cdot W_Q, \quad K = \Psi \cdot W_K, \quad V = \Psi \cdot W_V. \tag{3}$$

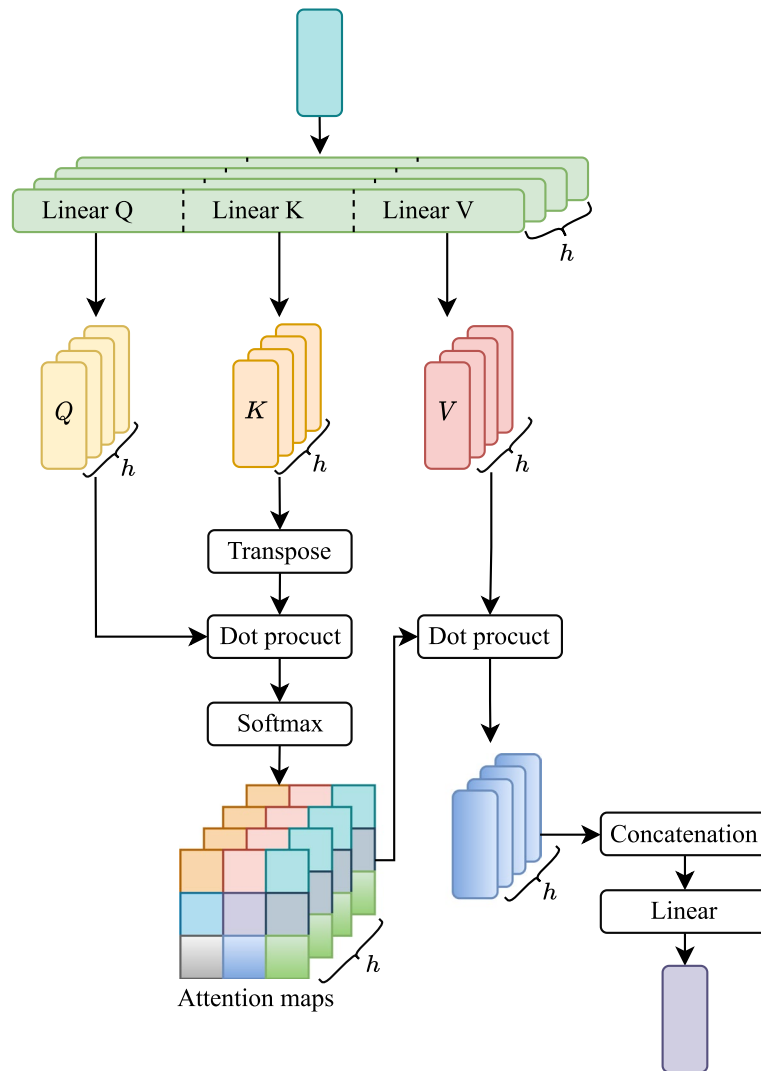


Fig. 5 Multi-head attention mechanism calculation process

Subsequently, each mapping is split into h independent attention heads, for $i = 1, \dots, h$, each head i utilizing distinct weight matrices $W_{Q_i}, W_{K_i}, W_{V_i}$.

Next, attention scores are computed for each head i , using the dot product of queries Q_i and keys K_i , normalized by the square root of the dimensionality d_k :

$$\psi_i = \text{Softmax}\left(\frac{Q_i \cdot K_i^T}{\sqrt{d_k}}\right) \cdot V_i, \quad (4)$$

where d_k is the dimensionality of Q_i and K_i . Applying the softmax operation to each head yields attention weights Softmax_i , which are then applied to the corresponding values V_i .

Finally, the outputs from all heads are concatenated or averaged to obtain the ultimate multi-head attention output:

$$\Omega = \text{Concat}([\psi_1; \psi_2; \dots; \psi_h]). \quad (5)$$

For guidance, we employ the output of the multi-head attention mechanism to individually modulate each slice-level feature map from sagittal, coronal, and axis. This process ensures that each plane is specifically adjusted based on the guidance derived from the 3D feature information. Specifically, For each 2D feature map from different anatomical planes (sagittal $F_{2D_{sag}}$, coronal $F_{2D_{cor}}$ and axis $F_{2D_{axi}}$), we use the attentional guidance signal Ω for weighting to form the guided 2D features:

$$\begin{aligned} F_{g_{sag}} &= F_{2D_{sag}} \odot \Omega, \\ F_{g_{cor}} &= F_{2D_{cor}} \odot \Omega, \\ F_{g_{axi}} &= F_{2D_{axi}} \odot \Omega, \end{aligned} \quad (6)$$

where $F_{2D_{sag}}, F_{2D_{cor}}, F_{2D_{axi}}$ denote the feature maps from the sagittal, coronal, and axis 2D slice-level networks respectively, and \odot represents element-wise multiplication.

After guiding each of these feature maps, we concatenate them to form the final integrated feature representation F_g :

$$F_g = \text{Concat}(F_{g_{sag}}, F_{g_{cor}}, F_{g_{axi}}). \quad (7)$$

This concatenated feature F_g offers a comprehensive view, encompassing enhanced 2D features from all three anatomical planes adeptly informed by the spatial information discerned from the 3D data. This nuanced application of the guidance linear and multi-head attention mechanism ensures that each anatomical direction is distinctly influenced by the 3D features, providing a robust and detailed basis for the diagnostic tasks.

2D slice-level network

Alzheimer's Disease is a neurodegenerative condition marked by the progressive deterioration of crucial brain regions. Notably, the hippocampus, essential for memory formation, is often among the first areas impacted, leading to memory loss. As AD advances, other cerebral cortex areas, such as the amygdala, which is responsible for emotion regulation, and the hypothalamus, which manages daily physiological activities, also degenerate. These changes are interlinked, each affecting the other, and are critical to understanding AD's holistic progression. To accurately detect subtle changes in these key brain regions in AD patients and understand how these alterations collectively influence brain function from multi-planar 2D slices, our 2D slice-level network combines 2D CNN and 2D Swin Transformer with attention feature fusion mechanism. The input to the 2D slice-level feature network consists of 40 automatically selected central slices from three anatomical planes: sagittal, coronal, and axial. Each plane is processed separately by its dedicated slice-level network, with 3D image features providing guidance to account for the unique characteristics of each orientation.

Residual module and advanced module

The 2D slice-level feature extract network is designed for sophisticated feature extraction in complex image datasets. The network initiates with a standard 2D convolutional layer for preliminary feature detection. This is followed by batch normalization and ReLU activation, which provide stability and introduce non-linearity. The core of the network comprises multiple residual modules. Each primary residual module within the network contains two convolutional layers, accompanied by batch normalization and ReLU activation. A shortcut connection is included in these modules to mitigate the issue of vanishing gradients. On the basic standard residual module, we incorporate AFF [17] to introduce both local and global attention mechanisms for refined feature extraction to form our advanced module. As depicted in Fig. 3, these advanced modules employ dual branches at varying scales to extract channel attention weights: one for global feature channel attention via global pooling, and the other for local feature channel attention via point-wise convolution. After attention extraction, the feature maps are fused based on these attention weights.

Swin Transformer module

For deeper feature processing and relationship capturing, the network incorporates the Swin Transformer module. This module's window-based attention mechanism is pivotal for detecting complex patterns and contextual information, significantly surpassing

the capabilities of traditional convolutional methods. To effectively address the computational demands of global self-attention designed in conventional Transformer modules, the Swin Transformer employs multi-head self attention (MSA) within confined windows. The module is configured in two distinct ways: the Window-based MSA (W-MSA) focuses on local window self-attention, while the Shifted Window-based MSA (SW-MSA) enhances the facilitation of information interaction across different windows.

As illustrated in Fig. 6, the Swin Transformer module significantly enhances feature correlation, resulting in strong correlations compared to the medium correlations observed without it. This demonstrates that the inclusion of the Swin Transformer improves the correlation between the captured multi-plane and multi-slice features, thereby enhancing the robustness and accuracy of our method.

Classification module

The guided features from the sagittal, coronal, and axis planes are concatenated and then passed through the fully connected layer. The softmax function applied to the final layer’s output provides the probability of the subject’s categorization into specific classes, such as Alzheimer’s Disease or Cognitively Normal (CN). For classification purposes, we employ the cross entropy loss function, L_p , which is formulated to be straightforward yet effective. The loss function is defined as:

$$L_p = -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C y_{nc} \log(p_{nc}), \tag{8}$$

where N represents the total number of subjects in the dataset, and C denotes the number of categories. y_{nc} is an indicator variable that is 1 if the true class for the n -th subject is c , and 0 otherwise. The term p_{nc} represents the predicted probability that the n -th subject belongs to class c , as outputted by the model.

Experiment & results

Dataset and preprocessing

The dataset utilized in this research is sourced from two publicly accessible databases: the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (<http://adni.loni.usc.edu/>) [18] and the Open Access Series of Imaging Studies (OASIS) (<https://sites.wustl.edu/oasisbrains/>) [19]. Specifically, this study employs the ADNI-1, ADNI-2, and OASIS-1 datasets, which consist of T1-weighted magnetic resonance imaging (MRI) brain scans. The ADNI sample includes 524 subjects, comprising 254 patients with AD and 270 CN individuals. The dataset is divided into 60% for training, 20% for validation, and 20% for testing. To evaluate the model’s generalizability, an independent test set consisting of 50 AD and 48 CN subjects from the OASIS dataset is used. The demographic characteristics of the subjects are detailed in Table 1. Additionally, cognitive function is assessed using the Mini-Mental State Examination (MMSE) scores, where higher scores indicate better cognitive performance and lower scores suggest potential cognitive impairment.

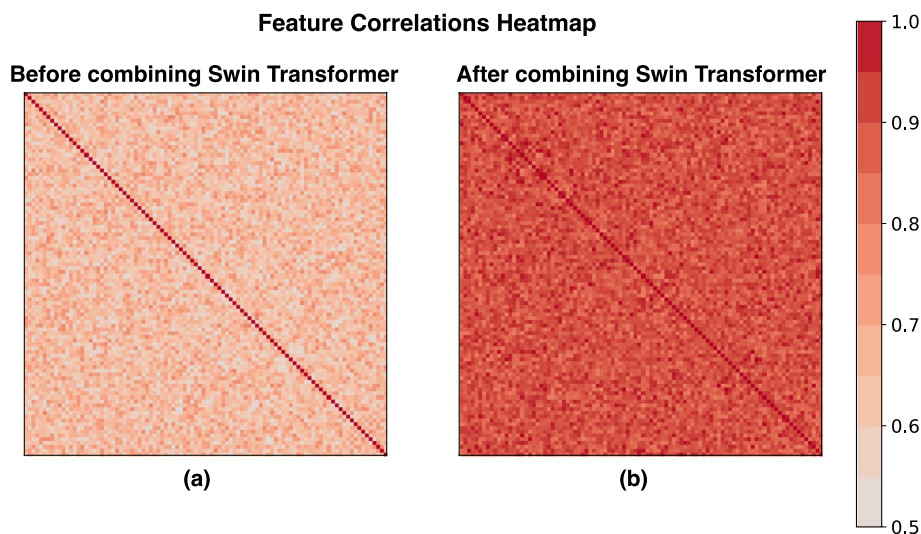


Fig. 6 Feature correlations heatmaps before and after combining with the Swin Transformer module. Values closer to 1 indicate a stronger correlation

Table 1 Demographic information for each dataset and category. The MMSE score ranges from 0 to 30, with higher scores indicating better cognitive function. Each subject in every group has one 3D image and 120 slices (40 coronal, 40 sagittal, and 40 axial)

Dataset	Group	Train (3D images/2D slices)	Val (3D images/2D slices)	Test (3D images/2D slices)	Age (Mean \pm SD)	Gender (M/F)	MMSE (Mean \pm SD)
ADNI [18]	AD	152/18240	51/6120	51/6120	75.52 \pm 9.58	116/138	19.53 \pm 4.47
	MCI	205/24600	68/8160	68/8160	76.25 \pm 9.58	167/174	22.85 \pm 6.33
	CN	162/19440	54/6480	54/6480	76.26 \pm 6.76	125/145	26.13 \pm 5.83
OASIS-1 [19]	AD	-	-	50/6000	74.22 \pm 8.71	21/29	20.15 \pm 4.83
	CN	-	-	48/5760	73.80 \pm 7.85	26/22	27.20 \pm 5.90

To optimize the model's focus on relevant brain structures, a series of preprocessing steps is applied to T1-weighted MRI images before model training. These steps were conducted using the Computational Anatomy Toolbox CAT12 (available at <https://neuro-jena.github.io/cat/>). Figure 7 depicts the preprocessing pipeline, which includes AC-PC correction, alignment with the MNI template, skull stripping, and segmentation into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). Consequently, this process yields standardized 3D GM images with dimensions of $121 \times 145 \times 121$ with a spatial resolution of $1.5 \times 1.5 \times 1.5 \text{mm}^3$, serving as the input for our model. For the slice-level network, 40 2D slices are automatically selected from the central region of each plane. This choice is driven by the fact that the central region of the brain typically harbors the most critical anatomical structures pertinent to AD, thereby providing a focused and relevant dataset for the analysis.

Experimental setup and evaluation criteria

The proposed model is implemented on Python 3.7.16 and Pytorch 1.10.0 with an Intel Core i5-12400F with 16 GB of RAM and an NVIDIA GeForce RTX 3090 GPU 24GB. The loss function in Eq. (8) is adopted to supervise the learning of the model parameters, which are optimized by the Adam optimizer with a learning rate of 0.001. To avoid over-fitting, we add an early stopping mechanism during the training process.

We evaluated the model performance from multiple perspectives by using metrics including classification accuracy (ACC), sensitivity (SEN), specificity (SPE), F1 score (F1), and the area under the receiver operating characteristic curve (AUC). These metrics are respectively defined as:

$$\begin{aligned}
 \text{ACC} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \\
 \text{SEN} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\
 \text{SPE} &= \frac{\text{TN}}{\text{TN} + \text{FP}}, \\
 \text{F1} &= \frac{2 \times \text{PRE} \times \text{SEN}}{\text{PRE} + \text{SEN}}, \\
 \text{PRE} &= \frac{\text{TP}}{\text{TP} + \text{FP}},
 \end{aligned} \tag{9}$$

where TP denotes true positive, TN denotes true negative, FP denotes false positive, and FN denotes false negative. The AUC characterizes the classification performance of the methods, the performance is better when AUC is closer to 1.

Comparison with different methods

To demonstrate the effectiveness of the proposed MHAGuideNet, we conducted a comparative analysis of the task of AD vs. CN classification with other methods. All methods are compared by training and testing on the same subjects from the ADNI dataset. Specifically, we have reproduced three models that utilize 3D sMRI: 3D Trans-ResNet [20], 3D ResNet [21], and 3D Swin Transformer [22]. Additionally, two 2D models are included that use the coronal slices from the same train dataset, with each subject's central 40 slices selected: DE-ViT [23] (based on Vision Transformer) and 2D ResNet [24]. Furthermore, we have also reproduced a combined 3D and 2D model, M3T [25], for greater comparison. The experiment results are shown in Table 2. The results presented in the table demonstrate the effectiveness of combining 3D images and 2D slices for AD vs. CN classification. The proposed MHAGuideNet outperforms models that utilize only 2D or 3D modalities, achieving an accuracy of 0.9758 and an AUC of 0.9931. Furthermore, the incorporation of transformer models enhances results compared to traditional CNN architectures, as evidenced by

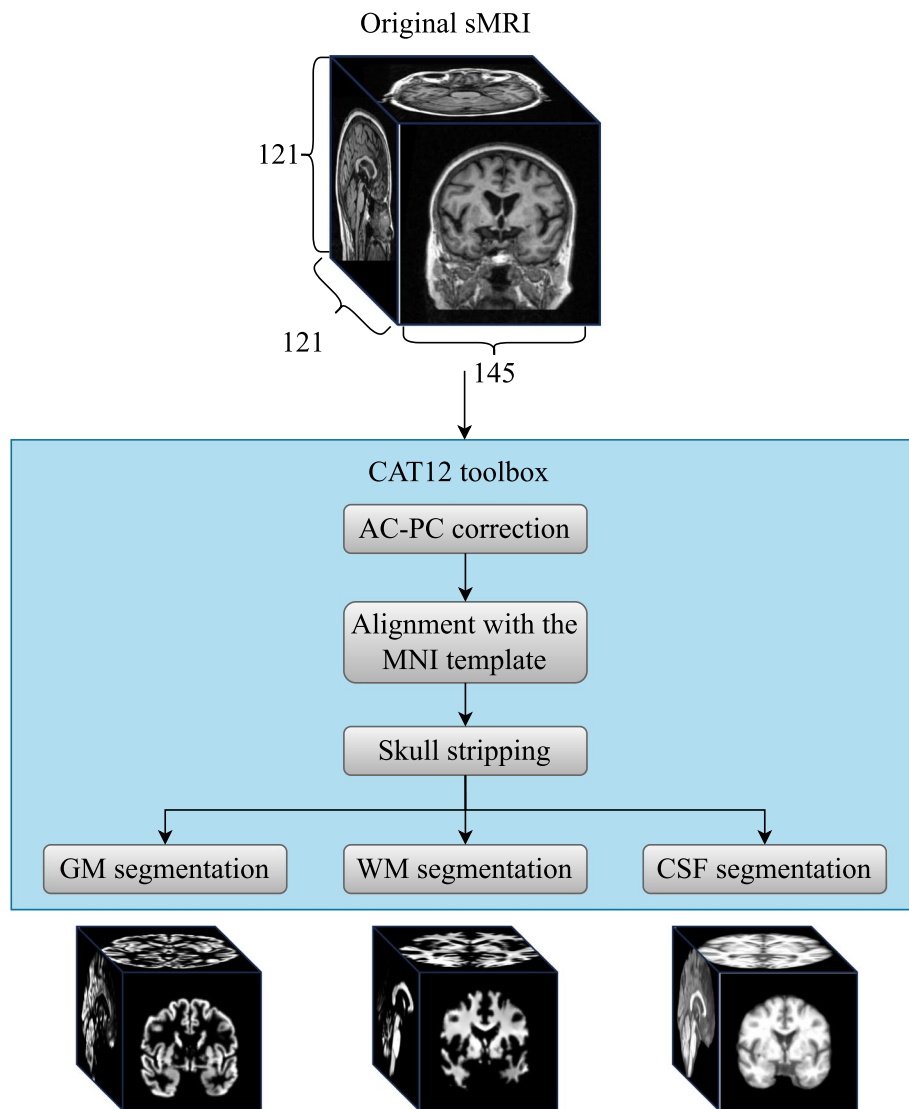


Fig. 7 The preprocessing pipeline of structural magnetic resonance imaging. The pipeline includes AC-PC correction, alignment with the MNI template, skull stripping and segmentation

Table 2 Comparison of different models for AD vs. CN classification, trained on the ADNI training dataset and tested on the ADNI test dataset

Method	Architecture	ADNI test dataset				
		ACC	SEN	SPE	F1	AUC
Only using 3D images						
3D Trans-ResNet [20]	3D CNN + 2D Transformer	0.9143	0.8431	0.9815	0.9053	0.9683
3D ResNet [21]	3D CNN	0.8095	0.7843	0.8333	0.8000	0.9187
3D Swin Transformer [22]	3D Transformer	0.8857	0.8431	0.9258	0.8776	0.9330
Only using 2D slices						
DE-ViT [23]	2D Transformer	0.9048	0.9412	0.9704	0.9057	0.9563
2D ResNet [24]	2D CNN	0.7905	0.8431	0.7407	0.7963	0.8228
3D images + 2D slices						
M3T [25]	3D CNN + 2D CNN + 2D Transformer	0.9616	0.9412	0.9815	0.9600	0.9899
MHAGuideNet (ours)	3D CNN + 2D CNN + 2D Transformer	0.9758	0.8863	0.9989	0.9398	0.9931

the higher accuracy of the 3D Trans-ResNet at 0.9143. In comparison with the M3T model, although our method falls slightly short in terms of sensitivity and F1 score, our model provides better overall accuracy, specificity, and AUC.

Ablation studies

Impact of guidance network

To emphasize the importance of the guidance network, we compare the proposed MHAGuideNet against models using the only 3D CNN network and the only 2D slice-level network. Furthermore, To highlight the critical role of the guidance network in processing multi-planer data (coronal, sagittal, and axial planes) for enhanced model performance, we examine two configurations: (1) first concatenating features from the three planes and then applying the guidance network on the combined features, and (2) applying the guidance network independently on each plane before concatenating them (as implemented in our MHAGuideNet). It is important to note that we train the models on the four configurations mentioned above using the ADNI dataset. The diagrams illustrating the training loss, validation loss, and AUC are presented in Fig. 8.

The results, as presented in Table 3, demonstrate that employing the guidance network individually on each plane leads to remarkable improvements compared to the 3D-only and 2D-only networks on both the ADNI and OASIS test datasets. On the ADNI test dataset, MHAGuideNet achieves a 2.90% increase in accuracy and 2.05% in AUC compared to the only 3D CNN network. Relative to the 2D-only network, MHAGuideNet achieves improvements of 1.45% in accuracy and 0.24% in AUC. These metrics highlight the model's superior diagnostic performance. Additionally, only the 3D CNN network and the 2D slice-level network exhibit a noticeable drop in performance on the independent OASIS test dataset, indicating limited generalization beyond the ADNI dataset. In contrast, MHAGuideNet maintains

robust performance across both datasets, highlighting its enhanced generalization ability. Moreover, the results reveal that independently processing each plane with the guidance network achieves improvements across performance metrics.

Impact of slice number

Table 4 presents a comparison of performance metrics between the 40-slice and full-slice configurations on the ADNI test dataset. The results show that both configurations perform comparably across most metrics, with only minor variations. Notably, the sensitivity is slightly higher in the full-slice configuration, potentially due to the broader range of diagnostic features captured by using all slices. However, this comes at the cost of a longer average prediction runtime (RT) of 0.8746 seconds, compared to 0.6532 seconds for the 40-slice configuration.

Overall, these results suggest that the full-slice setup offers a marginal increase in sensitivity but at the expense of efficiency, while the 40-slice configuration achieves similar performance with a more favorable runtime, making it a more practical choice for time-sensitive applications.

Impact of different planes

To gain a deeper understanding of the efficiency of our model in guiding multi-planar slices (coronal, sagittal, and axial planes), we conduct a comparative analysis of employing slice-level features within individual planes and across all planes. As shown in Table 5, the results reveal optimal performance metrics, including ACC, SEN, SPE, F1, and AUC, when considering information from all three planes simultaneously. This demonstrates the model's capacity to capture intricate features by integrating data from multiple planes. Compared with the sagittal plane and the coronal plane, the axial plane performs better on ACC, SEN, and SPE. Considering clinicians mainly analyze the ventricle enlargement in the axial or coronal planes, and hippocampus atrophy in the

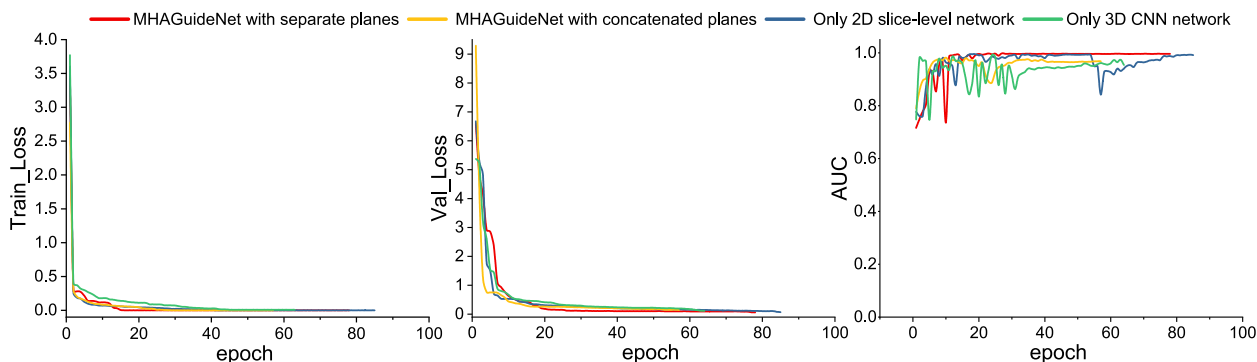


Fig. 8 Diagram illustrating the training loss, validation loss and AUC during the training on the ADNI dataset

Table 3 Comparison of classification performance with and without the guidance network on two datasets

Different configuration	ADNI test dataset					OASIS test dataset				
	ACC	SEN	SPE	F1	AUC	ACC	SEN	SPE	F1	AUC
Only 3DCNN network	0.9468	0.7500	0.9985	0.8571	0.9726	0.7500	0.6800	0.8261	0.7600	0.8313
Only 2D slice-level network	0.9613	0.8636	0.9877	0.9047	0.9907	0.7755	0.7000	0.8542	0.7609	0.8235
MHAGuideNet with concatenated planes	0.9614	0.8409	0.9938	0.9024	0.9732	0.9184	0.8642	0.9800	0.9245	0.9577
MHAGuideNet with separate planes	0.9758	0.8863	0.9989	0.9398	0.9931	0.9602	0.8720	0.9910	0.9271	0.9885
- Compare with 3D network only	(↑2.9%)	(↑13.63%)	(↑0.04%)	(↑8.72%)	(↑2.05%)	(↑21.02%)	(↑19.20%)	(↑16.49%)	(↑16.71%)	(↑15.72%)
- Compare with 2D network only	(↑1.45%)	(↑2.27%)	(↑1.12%)	(↑3.51%)	(↑0.24%)	(↑18.47%)	(↑17.20%)	(↑13.68%)	(↑16.62%)	(↑16.50%)

Table 4 Performance comparison between 40-slice and full-slice configurations on the ADNI test dataset. RT represents the average prediction runtime, with lower values indicating better performance

Slice Number	ACC	SEN	SPE	F1	AUC	RT (s)
Full Slices	0.9524	0.9216	0.9815	0.9495	0.9910	0.8746
40 Slices	0.9758	0.8863	0.9989	0.9398	0.9931	0.6532

Table 5 The experiment results of different planes used for guidance, including using sagittal plane, coronal plane, axial plane and multi planes

Different Plane	ACC	SEN	SPE	F1	AUC
Sagittal	0.9557	0.8324	0.9748	0.9175	0.9817
Coronal	0.9688	0.8465	0.9755	0.9078	0.9917
Axial	0.9691	0.8544	0.9877	0.9014	0.9879
ALL	0.9758	0.8863	0.9989	0.9398	0.9931

coronal plane [26, 27], our model MHAGuideNet has different abilities to analyze each plane and we can observe the importance to use all of the three planes in classifying sMRI images.

Impact of Swin Transformer module

The changes observed in AD subjects, such as cortical and hippocampal atrophy, are interconnected rather than isolated. The Swin Transformer module, integrated into our model, exhibits the capability to discern subtle alterations in these pivotal brain regions and capture interconnections between these changes. To assess the influence of incorporating the Swin Transformer in our model, as shown in the results presented in Table 6, we evaluate the impact of utilizing different quantities of Swin Transformer modules, ranging from zero to three, at the original spatial locations. The experimental results show that the model achieves the highest performance in terms of ACC, SEN, F1, and AUC when three Swin Transformer modules are utilized. However, considering the comprehensive factors such as computational complexity and model parameters, we find that the inclusion of only two Swin Transformer is sufficient.

Table 6 The impact of Swin Transformer module with different number

Num	ACC	SEN	SPE	F1	AUC
0	0.9302	0.8147	0.9212	0.8878	0.9407
1	0.9596	0.8814	0.9747	0.9270	0.9789
2	0.9758	0.8863	0.9989	0.9398	0.9931
3	0.9865	0.8977	0.9948	0.9504	0.9947

Robustness evaluation

To ensure the robustness of MHAGuideNet, we conduct 5-fold cross-validation on the ADNI datasets, using 203 AD subjects and 216 CN subjects in the train dataset and validation dataset, excluding those in the test dataset. As shown in Table 7, the mean accuracy, sensitivity, specificity, AUC, and F1 score across 5 folds are all high, with low standard deviations, indicating the model’s stability and robustness across different data splits.

Simultaneously, we simulate real-world imaging artifacts by adding Gaussian noise to the sMRI images in the ADNI datasets. As shown in Table 8, the model maintains high performance even as noise levels increase, with only a slight degradation in metrics, demonstrating its resilience to noisy inputs.

Attention maps

To demonstrate the ability of the proposed MHAGuideNet in extracting features, we employ Grad-CAM [28] for feature visualization. Figure 9 illustrates notable brain regions in attention areas in the AD and CN classification tasks from the sagittal, coronal, and axial views, respectively. In the task of diagnosis of AD, the model’s attention spans across areas of the cortex brain, with a pronounced focus on the ventricles and the hippocampus in three planes which indicates a widespread

Table 7 5-fold cross-validation of the proposed MHAGuideNet on the ADNI dataset

	ACC	SEN	SPE	F1	AUC
Mean	0.9689	0.8901	0.9930	0.9302	0.9926
Std Dev	0.0052	0.0087	0.0011	0.0043	0.0024

Table 8 Noise robustness validation of the proposed MHAGuideNet on the ADNI dataset

	ACC	SEN	SPE	F1	AUC
0% Noise	0.9758	0.8863	0.9989	0.9398	0.9931
10% Noise	0.9645	0.8734	0.9901	0.9287	0.9856
20% Noise	0.9532	0.8605	0.9813	0.9176	0.9781
30% Noise	0.9418	0.8476	0.9725	0.9065	0.9707

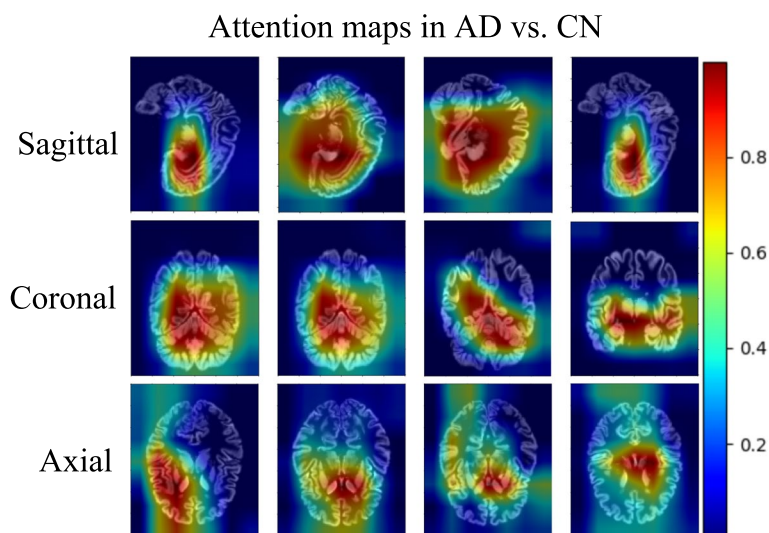


Fig. 9 The attention maps on sagittal, coronal, and axial plane in the AD vs. CN task

cortical involvement in AD. It is worth noting that these highlighted brain regions in the AD diagnosis align with findings from earlier clinical studies [29]. Additionally, Fig. 9 demonstrates that the heat map areas are distributed across three planes. This distribution suggests that our model has the potential ability to analyze AD-related abnormalities in the brain.

Discussion

Although many works have achieved encouraging results in computer-aided AD diagnosis based on sMRI using deep learning, there are still many limitations in applying the methods to the clinic. Traditional 3D CNN-based methods training can lead to increased model complexity and computational overhead, while 2D slice-based methods may miss important multi-planar information. In this study, we propose MHAGuideNet to address these limitations. Our method employs a guidance network to leverage 3D volumetric information for guiding 2D slice feature extraction across three planes, effectively capturing comprehensive spatial information with reduced computational cost. Additionally, the integration of CNN and Swin Transformer allows the model to capture the interrelations between different structural atrophies associated with Alzheimer’s Disease, improving the detection of structural changes associated with AD.

Experiment results reveal that integrating 3D volumetric guidance with multi-planar 2D slice-level feature extraction significantly improves diagnostic performance compared to using 3D or 2D data alone. By integrating CNN and Swin Transformer, our model effectively

captures both local structural features and long-range dependencies, addressing the limitations of CNNs in handling global context and structural relationships. Our comparisons suggest that MHAGuideNet achieves competitive performance relative to other methods, highlighting its effectiveness.

In the early stages of Alzheimer’s Disease, identifying and diagnosing mild cognitive impairment (MCI) is crucial for enabling early interventions. Consequently, we also conduct experiments to assess the performance of the proposed MHAGuideNet in the MCI diagnosis. We use 341 MCI subjects from the ADNI datasets, and the demographic information is provided in Table 1.

As shown in Fig. 10, we design two distinct binary classification tasks: AD vs. MCI and MCI vs. CN. The proposed MHAGuideNet demonstrates limitations in classifying MCI, achieving accuracies of only 74.41% for AD vs. MCI and 80.78% for MCI vs. CN. These results highlight the challenges in distinguishing between subtle differences in cognitive impairment stages, which may be due to overlapping characteristics among these categories. Addressing this issue may require further exploration of additional features.

Another limitation is observed when the model is tested on unseen OASIS datasets. Although the performance metrics remain relatively high, there is a slight decrease in accuracy compared to the results obtained on the ADNI dataset. This suggests that while the model generalizes well, it may not fully capture the variations present in different datasets, indicating room for improvement in its adaptability to new data.

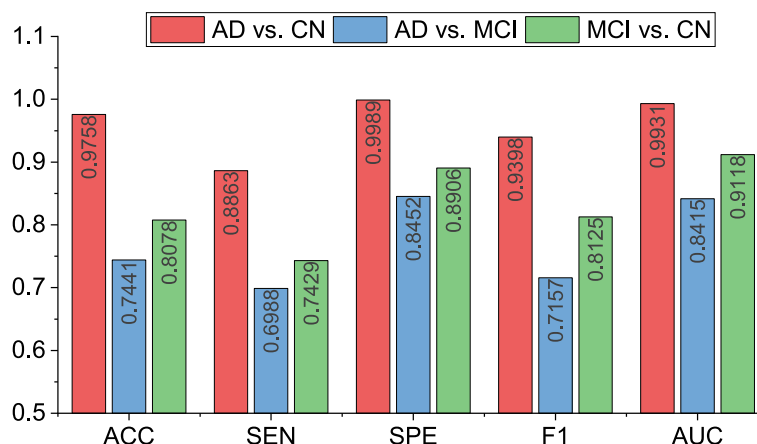


Fig. 10 The classification performance of the MHAGuideNet in different tasks: AD vs. CN, AD vs. MCI, MCI vs. CN

Conclusion

Alzheimer's Disease is a neurodegenerative condition that leads to irreversible and progressive brain damage, often characterized by structural atrophy. Computer-aided diagnostic methods based on sMRI data effectively identify these pathological features. However, existing 2D methods struggle to capture comprehensive multi-planar information, while 3D approaches are prone to overfitting and high computational overhead. To address these limitations, this study proposes MHAGuideNet, which leverages 3D guidance information and 2D slice features to enhance diagnostic accuracy and robustness. The study demonstrates that incorporating 3D guidance with multi-planar 2D slices and combining CNN with Swin Transformer enhances diagnostic performance, robustness, and the ability to capture both localized atrophy and spatial relationships.

Comprehensive evaluations using the ADNI test dataset demonstrate that MHAGuideNet achieves an accuracy of 97.58%, specificity of 99.89%, and AUC of 99.31% in the classification of AD versus CN subjects. Moreover, on the independent OASIS test dataset, the model maintains a robust performance with an accuracy of 96.02% and an AUC of 98.85%. Compared to using only the 3D CNN network, MHAGuideNet demonstrates an improvement of 2.90% in accuracy. When compared to using only the 2D slice-level network, the increase is 1.45% in accuracy. These results prove the effectiveness of the proposed model in fully utilizing spatial information from 3D images and local detail features.

Future research will focus on enhancing the model's ability to detect MCI, further advancing early diagnosis of Alzheimer's Disease. Additionally, incorporating clinical text data with imaging features to develop a more comprehensive diagnostic framework could contribute to developing a more comprehensive diagnostic framework.

Acknowledgements

We would like to thank the Alzheimer's Disease Neuroimaging Initiative (ADNI) and Open Access Series of Imaging Studies (OASIS) for providing access to the data required for this study.

Clinical trial number

Not applicable.

Authors' contributions

Y.N. implemented the study, drafted the manuscript, implemented the experiment, and interpreted the data. Q.C. helped design the classification network and contributed to the pertaining sections of the manuscript. W.L. contributed to the completion of the manuscript. Y.L. provided funding and helped revise the manuscript. T.D. evaluated the method. All authors read and approved the final manuscript.

Funding

This work was supported by the Science Innovation Programs Led by Academicians in Chongqing [grant numbers 2022yszx-jsx0002cstb]; and the Science and Technology Research Program of Chongqing Education Commission [grant numbers KJQN201900109].

Data availability

The sMRI image data collected are available as open data via the Alzheimer's Disease Neuroimaging Initiative and Open Access Series of Imaging Studies, the URLs are: <http://adni.loni.usc.edu/about/contact-us/> and <https://sites.wustl.edu/oasisbrains/>.

Declarations

Ethics approval and consent to participate

Ethics approval and consent to participate: As per ADNI protocols, all procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. More details can be found at adni.loni.usc.edu.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 28 August 2024 Accepted: 27 November 2024
Published online: 18 December 2024

References

- McDade E, Bateman RJ. Stop Alzheimer's before it starts. *Nature*. 2017;547(7662):153–5.
- Knopman DS, Amieva H, Petersen RC, Chételat G, Holtzman DM, Hyman BT, et al. Alzheimer disease. *Nat Rev Dis Prim*. 2021;7(1):33.
- Scheltens P, Leys D, Barkhof F, Huglo D, Weinstein H, Vermersch P, et al. Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. *J Neurol Neurosurg Psychiatry*. 1992;55(10):967–72.
- Bakkouri I, Afdel K. DermoNet: A computer-aided diagnosis system for dermoscopic disease recognition. In: *Image and Signal Processing: 9th International Conference, ICISP 2020, Marrakesh, Morocco, June 4–6, 2020, Proceedings 9*. Springer; 2020. pp. 170–177.
- Westman E, Aguilar C, Muehlboeck JS, Simmons A. Regional magnetic resonance imaging measures for multivariate analysis in Alzheimer's disease and mild cognitive impairment. *Brain Topogr*. 2013;26:9–23.
- Bloch L, Friedrich CM, Initiative ADN, et al. Systematic comparison of 3D Deep learning and classical machine learning explanations for Alzheimer's Disease detection. *Comput Biol Med*. 2024;170:108029.
- Mofrad SA, Lundervold A, Lundervold AS, Initiative ADN, et al. A predictive framework based on brain volume trajectories enabling early detection of Alzheimer's disease. *Comput Med Imaging Graph*. 2021;90:101910.
- Fard AS, Reutens DC, Vegh V. From CNNs to GANs for cross-modality medical image estimation. *Comput Biol Med*. 2022;146:105556.
- Basheera S, Ram MSS. A novel CNN based Alzheimer's disease classification using hybrid enhanced ICA segmented gray matter of MRI. *Comput Med Imaging Graph*. 2020;81:101713.
- Liu M, Zhang J, Adeli E, Shen D. Landmark-based deep multi-instance learning for brain disease diagnosis. *Med Image Anal*. 2018;43:157–68.
- Qiu S, Joshi PS, Miller MI, Xue C, Zhou X, Karjadi C, et al. Development and validation of an interpretable deep learning framework for Alzheimer's disease classification. *Brain*. 2020;143(6):1920–33.
- Bakkouri I, Bakkouri S. 2MGAS-Net: multi-level multi-scale gated attentional squeezed network for polyp segmentation. *SIVIP*. 2024;18:5377–86.
- Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: Transformers for image recognition at scale. 2020. arXiv preprint [arXiv:2010.11929](https://arxiv.org/abs/2010.11929).
- Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin transformer: Hierarchical vision transformer using shifted windows. In: *Proceedings of the IEEE/CVF international conference on computer vision*. Piscataway: IEEE; 2021. p. 10012–22.
- Xin J, Wang A, Guo R, Liu W, Tang X. CNN and swin-transformer based efficient model for Alzheimer's disease diagnosis with sMRI. *Biomed Signal Process Control*. 2023;86:105189.
- Xu X, Lin L, Sun S, Wu S. A review of the application of three-dimensional convolutional neural networks for the diagnosis of Alzheimer's disease using neuroimaging. *Rev Neurosci*. 2023;34(6):649–70.
- Dai Y, Gieseke F, Oehmcke S, Wu Y, Barnard K. Attentional feature fusion. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. Piscataway: IEEE; 2021. p. 3560–9.
- Jack CR Jr, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J Magn Reson Imaging Off J Int Soc Magn Reson Med*. 2008;27(4):685–91.
- Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J Cogn Neurosci*. 2007;19(9):1498–507.
- Li C, Cui Y, Luo N, Liu Y, Bourgeat P, Fripp J, et al. Trans-resnet: integrating transformers and cnns for alzheimer's disease classification. In: *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*. IEEE; 2022. pp. 1–5.
- Korolev S, Safiullin A, Belyaev M, Residual Dodonova Y, plain convolutional neural networks for 3D brain MRI classification. In: *2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017)*. IEEE; 2017. p. 835–8.
- Tang Y, Yang D, Li W, Roth HR, Landman B, Xu D, et al. Self-supervised pre-training of swin transformers for 3d medical image analysis. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. Piscataway: IEEE; 2022. p. 20730–40.
- Sen A, Roy S, Debnath A, Jha G, Ghosh R. DE-ViT: State-Of-The-Art Vision Transformer Model for Early Detection of Alzheimer's Disease. In: *2024 National Conference on Communications (NCC)*. IEEE; 2024. pp. 1–6.
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. Piscataway: IEEE; 2016. p. 770–8.
- Jang J, Hwang D. M3T: three-dimensional Medical image classifier using Multi-plane and Multi-slice Transformer. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. Piscataway: IEEE; 2022. p. 20718–29.
- Scheltens P, Pasquier F, Weerts JG, Barkhof F, Leys D. Qualitative assessment of cerebral atrophy on MRI: inter-and intra-observer reproducibility in dementia and normal aging. *Eur Neurol*. 1997;37(2):95–9.
- Scheltens P, Launer LJ, Barkhof F, Weinstein HC, Van Gool WA. Visual assessment of medial temporal lobe atrophy on magnetic resonance imaging: interobserver reliability. *J Neurol*. 1995;242:557–60.
- Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE international conference on computer vision*. Piscataway: IEEE; 2017. p. 618–26.
- Galton CJ, Patterson K, Graham K, Lambon-Ralph MA, Williams G, Antoun N, et al. Differing patterns of temporal atrophy in Alzheimer's disease and semantic dementia. *Neurology*. 2001;57(2):216–25.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.