Baghdad Science Journal

Volume 23 | Issue 5 Article 16

5-26-2026

Cross-Scale Feature Learning for Alzheimer's Disease Detection via Hybrid Deep Networks

Essam Haidar Majeed
Department of Computer Systems Techniques, Technical Institute of Najaf, Al-Furat Al-Awsat Technical University, Najaf, 54001, Iraq, essam.mageed.cku@atu.edu.iq

Fallah H. Najjar
Department of Computer Systems Techniques, Technical Institute of Najaf, Al-Furat Al-Awsat Technical University, Najaf, 54001, Iraq AND Department of Emergent Computing, Faculty of Computing, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia, fallahnajjar@atu.edu.iq

Safaa M. Kareem
Department of Computer Systems Techniques, Technical Institute of Najaf, Al-Furat Al-Awsat Technical University, Najaf, 54001, Iraq, safaa.kareem@atu.edu.iq

Munaf H. Kareem
Department of Computer Systems Techniques, Technical Institute of Najaf, Al-Furat Al-Awsat Technical University, Najaf, 54001, Iraq, munaf.kareem@atu.edu.iq

Duha Amer Mahdi
Department of Computer Systems Techniques, Technical Institute of Najaf, Al-Furat Al-Awsat Technical University, Najaf, 54001, Iraq, duha.amer@atu.edu.iq

How to Cite this Article

Majeed, Essam Haidar; Najjar, Fallah H.; Kareem, Safaa M.; Kareem, Munaf H.; and Mahdi, Duha Amer (2026) "Cross-Scale Feature Learning for Alzheimer's Disease Detection via Hybrid Deep Networks," *Baghdad Science Journal*: Vol. 23: Iss. 5, Article 16.
DOI: https://doi.org/10.21123/2411-7986.5300

BAGHDAD SCIENCE JOURNAL 2026;23(5):1779–1794

Scan the QR to view the full-text article on the journal website

RESEARCH ARTICLE

Cross-Scale Feature Learning for Alzheimer's Disease Detection via Hybrid Deep Networks

Essam Haidar Majeed [1], **Fallah H. Najjar** [1,2,*], **Safaa M. Kareem** [1], **Munaf H. Kareem** [1], **Duha Amer Mahdi** [1]

[1] Department of Computer Systems Techniques, Technical Institute of Najaf, Al-Furat Al-Awsat Technical University, Najaf, 54001, Iraq
[2] Department of Emergent Computing, Faculty of Computing, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

ABSTRACT

Alzheimer's disease (AD) is a chronic neurodegenerative disorder that is the leading cause of dementia. Early diagnosis is essential for interepting the progression of cognitive decline and favoring the clinical outcome, although conventional diagnostic means may not be able to identify early symptomatic forms. In this manuscript, we propose a residual attention-based hybrid deep learning model with the EfficientNetB1 backbone for improving the classification of AD based on structural MRI data obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and ADNI1 datasets. The network structure integrates two-dimensional (2D) convolutional feature extraction, multi-head self-attention for building the global context information, and residual learning. The proposed method can acquire both the local anatomical characters and the global brain structure dependencies effectively. It outperforms several state-of-the-art methods in terms of diagnostic accuracy on both models, reaching macro-averaged accuracy of 99.13% and 99.37% over ADNI and ADNI1 data, respectively, based on three anatomical planes (axial, sagittal, and coronal). This then led to biologically feasible areas from attention maps, which enhanced interpretability. These findings indicate that our model is an effective and interpretable approach to early and accurate classification of AD, offering the potential to be deployed in clinical settings. In future, the model will be evaluated for its usability and pragmatic effectiveness in clinical use to confirm sufficient applicability for neurologists and radiologists.

Keywords: Alzheimer's disease, Cross-scale feature, Classification, Deep learning, Hybrid model

Introduction

Alzheimer's disease (AD) is a neurodegenerative condition marked by progressive loss of cognitive function, memory disturbance, and alterations in behavior.[1] AD is the leading cause of dementia and represents a major public health issue affecting millions of people and their families across the globe.[2] Recent reports suggest that about 50 million individuals worldwide are suffering from dementia, and AD is considered responsible for up to 70% of the cases.[3] With the increasingly aging population, it is estimated that the burden of the disease will increase significantly in the coming years, highlighting the importance of early detection and intervention.[4] Early prevention of AD is important for slowing disease progression, improving patients outcomes, and enabaling timely therapeutic intervention.

Nevertheless, conventional clinical diagnostics, including neuropsychological tests, biomarker measurements, and imaging techniques, primarily recognize the disease in a later stage.[5] Magnetic Resonance Imaging (MRI) has been proven to be a powerful, non-invasive approach to studying structural changes associated with AD, including hippocampal atrophy and cortical thinning.[6] However, the visual interpretation of MRI is time-consuming, has inter-rater variability, and is limited by the subtlest early disease manifestations, underscoring the need for computer-aided diagnosis.[7] On the other hand, the Alzheimer's

Received 16 June 2025; revised 26 December 2025; accepted 2 January 2026.
Available online 26 May 2026

* Corresponding author.
E-mail addresses: essam.mageed.cku@atu.edu.iq (E. H. Majeed), fallahnajjar@atu.edu.iq (F. H. Najjar), safaa.kareem@atu.edu.iq (S. M. Kareem), munaf.kareem@atu.edu.iq (M. H. Kareem), duha.amer@atu.edu.iq (D. A. Mahdi).

https://doi.org/10.21123/2411-7986.5300

Disease Neuroimaging Initiative (ADNI) dataset provides an abundant and standardized data source of longitudinal MRI scans and clinical information, enabling the development of reliable, automated instruments to identify AD-related patterns more sensitively.[8] Meanwhile, Artificial Intelligence (AI) and Deep Learning, especially Convolutional Neural Networks (CNNs), have performed extraordinarily in computer-aided analysis of medical images, such as brain MRIs, in recent years.[9] CNNs can learn hierarchical features only from imaging data, making them suitable for distinguishing between healthy and AD classifications.[10] However, conventional CNNs can only locally leverage spatial information with a limited receptive field of convolutional operations, which may be insufficient to model the global structural changes of the brain during the development of AD. To alleviate these shortcomings, attention is embedded in deep learning architectures, enabling the models to focus on related parts of the input.[11] Additionally, residual learning, as in Residual Networks (ResNets), has been very useful for training deep models by creating shortcuts for gradient flow and allowing the networks to learn more complex mappings. It is an interesting approach to enhance the performance of deep learning models in challenging classification tasks, such as the automated diagnosis of AD from MRI scans.[12]

Despite the success of deep learning techniques in AD classification, most models struggle to learn both local anatomical and global structural information from MRI data.[13] In traditional CNN designs, the characteristics are incapable of dynamically selecting the localized region where relevant information is preserved, and the network becomes unstable without an effective optimization method when the network depth is too deep.[14] Thus, a stronger, interpretable, and robust framework is highly demanded to effectively integrate fine-grained local information and global context awareness as much as appropriate while mitigating the corresponding noise and remaining stable during the training of deep networks.[15]

The main objective of this work is to develop a Residual Attention Network that is effective at both the processing and feature levels, specifically for AD classification using MRI. The proposed model aims to leverage the benefits of convolutional feature extraction, attention-based focus on salient regions, and residual connections for smooth optimization. The goals are, more specifically: (i) To develop a deep learning network incorporating residual learning and attention mechanisms to enhance the classification accuracy of AD on the ADNI and ADNI1 datasets. (ii) To develop an explainable framework in which the attention maps of the network suggest biologically plausible regions related to the disease progression. By achieving these aims, our proposed model is designed to provide automated, accurate, and clinically interpretable tools for the early detection and diagnosis of AD.

Our main contributions can be summarized as follows: (i) We propose a novel deep learning architecture that combines the EfficientNetB1, residual bottleneck blocks, and multi-head self-attention to learn local anatomical structures and long-range structural relationships concurrently. (ii) Unlike the majority of prior publications, which perform classification using random subject-level splits, we follow the three anatomical planes — axial, sagittal, and coronal — to ensure validation of clinical generalization across multiple orientations. (iii) The attention score maps are biologically interpretable, which means the model uncovers which anatomical regions are recognized to make a decision, thereby promoting transparency and generating clinical trust.

We choose EfficientNet as a backbone because it efficiently uses the compound scaling method to scale network depth, width and resolution uniformly. Compared to ResNets, which scale only in depth, and DenseNet, which offers dense cross layers connectivity, EfficientNet provides much superior accuracy and efficiency trade-offs on image classification tasks. This is critical in medical imaging, where high-resolution 2D slices have to be processed in large quantities, and computational efficiency becomes a key factor. In particular, the EfficientNetB1 architecture is rich enough to capture detailed anatomical features while reamaining computationally affordable. This balance makes it particularly well-suited to structural MRI (sMRI) analysis in which both global contextual structures (e.g., ventricles) and local fine-grained structure (e.g., hippocampal atrophy) need to be identified.

To the best of our knowledge, our study is the first one that incorporates bottleneck residual units into a backbone of EfficientNetB1 and is coupled with cross-scale multi-head self-attention for AD classification from sMRI. It enables the network to learn fine-grained local anatomical features through residual bottlenecks, and long-range context between different resolutions through self-attention at the same time. In addition, contrary to previous efforts, which commonly perform slice-based or volume-based testing, we apply subject-level splitting and consider three anatomical planes (axial, coronal, and sagittal) for inclusive and clinical performance.

Related works

In recent years, significant advancements have been made in the classification and diagnosis of AD using

deep learning. Considrerable research has focused on developing models that can effectively accommodate the complexity of neuroimaging data, particularly sMRI, fMRI, Dermoscopic, and PET, while also being accurate, generalizable, and interpretable. [16]

Traditional CNN-based architectures remains ubiquitous in AD detection. Transfer learning methods have effectively utilized pre-trained networks, such as VGG19, InceptionV3, and ResNet50, as demonstrated by Tamanna et al., [17] resulting in an accuracy of 90% on the ADNI dataset with a limited number of training epochs. Similarly, Rani et al. [18] introduced a SqueezeNet with adaptive reptile search clustering based on features, achieving a classification accuracy of 98% by concentrating on a small footprint and efficient feature selection.

Residual networks are allowed to be deep without suffering from vanishing gradients. Hassan et al. [19] presented a five-stage residual-based multi-deep feature learning framework combining CNN with feature selection and classifiers such as SVM or Random Forest. Their proposed method achieved over 99% accuracy on three datasets (ADNI1, MIRAID, and OASIS) and demonstrated superior performance compared to state-of-the-art approaches. Kumar et al. [20] proposed a 3D wavelet-based method utilizing the features of hippocampal substructures with an attention mechanism, achieving 93.39% accuracy on the EADC-ADNI dataset.

More recently, the transformer has become incredibly promising. [21] Vision Transformers (ViTs), which were primarily developed for natural language processing, can be transferred to the neuro-image domain. Alp et al. [22] proposed a dual transformer architecture for 3D MRI classification, which extracted long-range dependencies among brain structures and achieved an accuracy of 99.04%. Similarly, Akan et al. employed a transformer in conjunction with bidirectional long short-term memory (Bi-LSTM) to effectively learn temporal and spatial information in binary classification. [23]

Attention-based approaches for improving interpretability and slice-level analysis have shown great potential. The AXIAL by Lozupone et al. [24] applied an attention-based mechanism to localize diagnostically important MRI slices, demonstrating better classification performance and model interpretability. HybridViT combined CNN backbones with transformer encoders to enhance multimodal representation learning for efficient multi-class AD diagnosis. [25]

Feature fusion has been a key focus in multimodal learning. The MFIFN model introduced by Huang et al., [26] employs a hybrid attention mechanism to integrate sMRI and fMRI. The proposed method incorporated CNN and GRU layers for temporal filtering effect and attained a classification accuracy of 99.3% on the ADNI dataset. Similarly, Huang et al. achieved 97.71% multi-class accuracy.

Multimodal and multi-scale integration is another growing area. Abdelaziz et al. [27] developed a multi-scale multimodal deep learning framework that utilized both MRI and PET data. Their architecture featured multi-head self-attention and cross-attention mechanisms to enhance inter-modality interactions across scales. The model outperformed existing patch-based methods on ADNI data by capturing finer anatomical and metabolic cues.

Li et al. [28] addressed the challenge of protocol inconsistency in MRI scans through a unified multi-protocol approach. Their method combined a dual-decoder adversarial autoencoder with an Ensemble Residual Shrinkage Attention Network (ERS2TA), aligning images from various MRI protocols and focusing on disease-relevant brain regions. Their system improved generalizability and diagnostic performance in clinically realistic settings.

Wavelet- and texture-based techniques also still have merits. Garg et al. [29] proposed a WSELBP to extract both global and local features of brain anatomy. Integrating spatial frequency information, this approach achieved over 95% classification accuracy, especially for early-stage diagnosis.

Genetic information integration and optimization-based models provide additional diagnostic approaches. Xin et al. [30] introduced an MDBO-SVM framework that integrates imaging genetics and a multi-strategy Dung Beetle Optimizer. This hybrid model achieved 81.7% training accuracy in multi-class classification, demonstrating the effectiveness of evolutionary optimization in AD modeling. Similarly, Alinsaif [31] integrated shearlet transform features with Discriminant Correlation Analysis (DCA) and deep learning to construct a DCA-enhanced AD detection system, achieving an accuracy exceeding 95%.

An alternative approach for addressing this is to utilize cross-modality methods. Liu et al. [32] proposed a two-stream CNN that employs progressive variation through spatial feature enhancement and dimensionality reduction to address spatial redundancy and enhance the model robustness.

Representation learning have also advanced considerably. The FiboNeXt model, named after the Fibonacci sequence, was introduced as a trade-off between model complexity and performance. [33] It processed high-dimensional neuroimaging data efficiently while maintaining spatial hierarchies through its CNN-based architecture. In summary, AD classification has evolved from basic CNN models to advanced multimodal and transformer-based ones. Attention mechanisms, multimodal feature

Table 1. Distribution of ADNI and ADNI1 datasets images.

Dataset	Class	Axial	Sagittal	Coronal	Total image per class
ADNI	CN	3200	–	–	3200
	MCI	2240	–	–	2240
	MC	896	–	–	896
	AD	64	–	–	64
	Total number	6400	–	–	6400
ADNI1	MCI	1113	1113	1113	3339
	CN	705	705	705	2115
	AD	476	476	476	1428
	Total number	2294	2294	2294	6882

interaction, protocol alignment, and genetic data fusion represent the major focus of recent studies. These improvements enhance not only the classification accuracy but also some real-world diagnostic problems, such as imaging protocols variability and model decisions interpretability.

Methodology

In this section, we introduce the design and building blocks of the proposed model for classifying the ADNI and ADNI1 datasets. We describes preprocessing the dataset, components of the architecture, mathematical formulae for the attention mechanism, and the training process.

Data acquisition

The datasets employed in this work comprise ADNI and ADNI1. The ADNI1_Complete_1Yr_1.5T datasets[34] were designated to facilitate research into the early detection and progression of AD. This dataset comprises T1-weighted MRI images acquired from 1.5 Tesla scanners and features a full 1-year follow-up, ensuring temporal consistency for longitudinal follow-up studies. Each subject's scan is represented in three anatomical planes (sagittal, axial, and coronal), allowing the extraction of spatial features in multiple orientations.

The ADNI1 dataset consists of 1113 Mild Cognitive Impairment (MCI) cases, 705 Cognitively Normal (CN) cases, and 469 AD cases for each anatomical view, with balanced representation across modalities. All three views used for each subject facilitate advanced deep learning models based on multi-view fusion or cross-scale feature learning, which can enhance classification accuracy among CN, MCI, and AD. This extensive dataset is ideal for developing robust diagnostic systems and investigating patterns of neurodegenerative progression. On the other hand, the ADNI (Kaggle dataset) consists of 896 Mild Cognitive (MC) cases, 3200 CN cases, 2240 MCI cases, and 64 AD cases. The distribution of the ADNI and ADNI1 datasets' images is listed in Table 1.

Unlike the ADNI1 dataset, which is based on standard diagnostic criteria (CN, MCI, and AD), the Kaggle's ADNI dataset uses severity-based diagnostic labels. In this dataset, VeryMildDemented is equivalent to MCI, and MildDemented denotes MC. As MC is not an established neuro-pathological stage within the ADNI diagnostic framework, this category is considered data-specific and excluded from statistical analysis along with MCI.

Data augmentation

We applied data augmentation during training to alleviate class imbalance and to help with generalization. Every input slice was augmented with random small in-plane rotations ($\pm 10^{\circ}$), horizontal and vertical flips, addition of Gaussian noise ($\sigma \leq 0.01$), and intensity scaling ($\pm 10\%$). Augmentation was utilized in a higher ratio for minority-class samples (especially AD) to equalize the effective training distribution.

Proposed method

The proposed classification method comprises several key components: a feature extraction backbone, a sequence reshaping head, a self-attention block, a residual learning block, global average pooling, and a fully connected classification head. These parts are combined to form a joint architecture that aims to capture the local and global features of MRI scans well. In the following, we describe each element with its clear roles in the system.

Our approach is based on the concept that Alzheimer's disease is characterized by both local anatomical changes (e.g., hippocampus atrophy, cortical thinning) and global structure alteration (e.g., disconnection between distant regions). Traditional CNNs are good at modeling fine-grained local features owing to a small receptive field; however, they suffer from long-range dependence modeling. In comparison, self-attention mechanisms are good at encoding

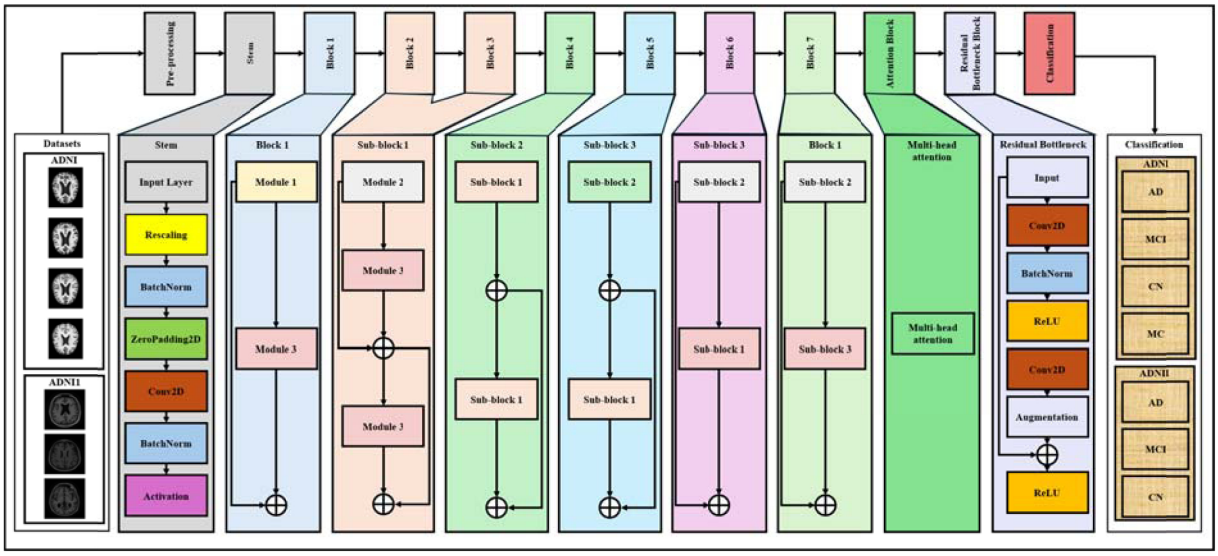

Fig. 1. Complete pipeline of the proposed model for AD classification.

global context information in that they can allow any spatial token to attend to any distant token. By combining these two paradigms via cross-scale feature learning, the model is designed to theoretically learn local details and global structures simultaneously. The residual bottleneck blocks also keep optimization simple and provide identity mappings, which ease the stacking of many layers to increase network depth. Collectively, these elements introduce a principled structure for capturing the multi-scale nature of AD-related brain changes that warrants architecture beyond empirical gains.

To effectively distinguish AD from MRI images, we employ a powerful convolutional feature extractor with global context modeling and deep feature refinement prior to the classification stage. The overall network structure is derived from EfficientNetB1 as its network backbone for feature extraction. It includes a Multi-Head Self-Attention module and the Residual Bottleneck block, enabling the network to store and process more information. After the convolution operation by EfficientNetB1, the feature maps are reshaped and fed into an attention model to compute global dependencies between regions. Figs. 1 and 2 flow into one another in an architectural sense in the model.

The modular components in Fig. 2 (i.e., multi-head attention and convolutional blocks) are shared across the larger architecture of Fig. 1, which enables the simultaneous extraction of local features and modeling of global contexts. A novel feature of this hybrid model is that low-level spatial encoding convolutional operations work in conjunction with attention-derived transformer operations for long-range feature association, enabling the network to process brain MRI images and conduct accurate classification based on different cognitive states.

CNNs are widely used in computer vision because they can efficiently represent local spatial hierarchies through localized computation. CNNs learn increasingly abstract feature representations at deeper layers by convolving filters on proximate pixels. Nevertheless, traditional approaches to scaling out CNNs, such as increasing network depth, width, or input resolution, are inefficient and may lead to overfitting. However, EfficientNetB1 solves this to some extent by scaling up all dimensions uniformly based on the same single coefficient ϕ via a principled method called compound scaling. The scaling relations are shown in Eq. (1):

$$depth = \alpha^{\phi}, width = \beta^{\phi}, resolution = \gamma^{\phi}, \text{subject to } \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \quad (1)$$

where ϕ represents the scaling factor that controls model size, α^{ϕ}, β^{ϕ}, γ^{ϕ} represent the constants that balance model dimensions,

Given an input image $I \in \mathbb{R}^{224\times224\times3}$, the EfficientNetB1 backbone produces feature maps Γ_0, as shown in Eq. (2):

$$\Gamma_0 = EfficientNetB1\,(I) \text{ where } \Gamma_0 \in \mathbb{R}^{H\times W\times C} \quad (2)$$

where H and W represent the reduced spatial dimensions, and C is the number of feature channels.

Although CNNs effectively capture local features, they struggle to capture global dependencies between distant input areas due to their limited receptive field

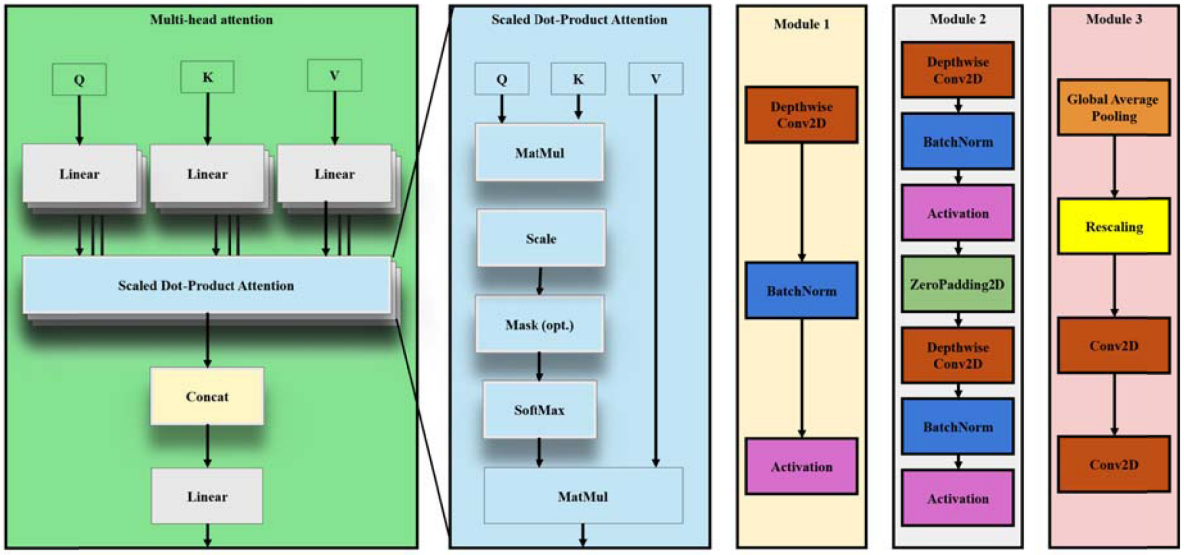

Fig. 2. Core modules of the proposed hybrid architecture.

size. To overcome this issue, the spatial feature maps need to be reshaped into sequences that can be used for attention. This reshaping transforms the feature map Γ_0 into a sequence of tokens using Eq. (3):

$$\Gamma_0' \in \mathbb{R}^{N\times C} \text{ where } N = H \times W \tag{3}$$

Each token represents a local portion of the original image that has been finely grained into an associated node with feature dimensionality C, allowing the feature set to conform with transformer-based models.

The self-attention mechanism can encode global context by enabling all tokens to freely attend to every other token, independent of their spatial distance. It results in a large receptive field around the image in a single layer, outperforming those convolutional layers that only capture information locally. Each token sequence is projected into Query Q, Key K, and Value V matrices, which are defined by Eq. (4):

$$Q = \Gamma_0' W_Q,\ K = \Gamma_0' W_K,\ V = \Gamma_0' W_V \quad \text{where } W_Q W_K W_V \in \mathbb{R}^{C\times d} \tag{4}$$

where d represents the dimensionality of each projection.

Eq. (5) is used to calculate the attention output for each query Q.

$$Attention\ (Q,\ K,\ V) = SoftMax\left(\frac{QK^T}{\sqrt{d}}\right) \times V \tag{5}$$

The scaling of $\sqrt{d}$ guarantees that no excessively large dot products will generate very small gradients for the softmax function, thereby stabilizing learning.

The Multi-Head Self-Attention (MHSA) module comprises multiple parallel attention operations that can be defined by Eq. (6).

$$MHSA\left(\Gamma_0'\right) = Concat\left(head_1, \ldots, head_h\right) \times W_0 \quad \text{where } W_0 \in \mathbb{R}^{hd\times C} \tag{6}$$

where h represents the head numbers, and d denotes the learned output projection.

The output following attention computation is fed into the position-wise Feed-Forward Network (FFN). The FFN introduces extra non-linearity to the transformation of token representations.

The FFN is composed of two fully connected layers, which can be calculated using Eq. (7):

$$FFN\,(x) = \sigma\left(xW_1 + b_1\right)W_2 + b_2 \quad \text{where } W_1 \in \mathbb{R}^{C\times 4C}, W_2 \in \mathbb{R}^{4C\times C} \tag{7}$$

where σ represents the Rectified Activation function (ReLU).

For stabilizing training and maintaining information flow, the attention output and the FFN output are wrapped with residual connections and layer normalization, as shown in Eqs. (8) and (9), respectively.

$$z = LayerNorm\,(x + MHSA\,(x)) \tag{8}$$

$$z = LayerNorm\left(y + FFN\left(y\right)\right) \tag{9}$$

Residual connections help to learn identity operations and alleviate the vanishing gradient problem of deep networks.

Since the downstream modules (e.g., Convolutional layers) are assuming a 4D tensor, i.e (batch size, height, width, and channels), the 2D sequence z needs to be transformed back into a 3D spatial feature map using Eq. (10):

$$z' \in \mathbb{R}^{H\times W\times C} \tag{10}$$

Deep architectures often suffer from the degradation phenomenon because an increased number of layers results in a higher training error. Residual blocks, proposed in ResNet, deal with this problem by learning residual functions $F(x)$ rather than directly mapping functions $H(x)$, so that the network can be trained more deeply without degrading its performance. The sequence of operations for the bottleneck residual block is calculated by Eqs. (11) to (14), respectively.

$$x_1 = \sigma\left(BN\left(Conv_{1\times1}\left(z', C/4\right)\right)\right) \tag{11}$$

$$x_2 = \sigma\left(BN\left(Conv_{3\times3}\left(x_1, C/4\right)\right)\right) \tag{12}$$

$$x_3 = BN\left(Conv_{3\times3}\left(x_2, C/4\right)\right) \tag{13}$$

$$x_4 = BN\left(Conv_{1\times1}\left(x_3, C\right)\right) \tag{14}$$

If the input and the output dimensions differ, a 1×1 projection is applied using Eq. (15).

$$Res = BN\left(Conv_{1\times1}\left(z', C\right)\right) \tag{15}$$

Finally, the residual addition is calculated by Eq. (16).

$$F = \sigma\left(x_4 + Res\right) \tag{16}$$

After residual learning, we use Global Average Pooling (GAP) to compress the 2D feature map into a single vector. For a feature map $F \in \mathbb{R}^{H\times W\times C}$, GAP for each channel c is calculated by Eq. (17):

$$g_c = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} F_{i,j,c} \tag{17}$$

The final vector is $g \in \mathbb{R}^C$, which significantly decreases the number of parameters compared to flattening and thereby improving generalization.

Following GAP, the low-dimensional feature vector is fed through fully connected layers, see Eq. (18), which are ultimately used to form nonlinear combinations of features and generate a probability distribution over the output classes.

$$h = \sigma\left(gW_3 + b_3\right) \text{ where } W_3 \in \mathbb{R}^{C\times256} \tag{18}$$

The dropout is applied before the final classification layer to prevent overfitting. Dropout randomly sets a proportion of input units to 0 at each training update, which helps prevent overfitting and allows the network to learn representations reminiscent of ensembles of networks, as presented using Eq. (19)

$$Z = hW_4 + b_4 \tag{19}$$

The final output layer is equipped with a SoftMax activation to guarantee the output is positive and sums to 1, which can thus be interpreted as class ratios using Eq. (20).

$$\hat{y}_i = \frac{e^{z_i}}{\sum_j e^{z_j}} \tag{20}$$

The model is trained using the categorical cross-entropy loss, a well-known loss function for multi-class classification tasks, as defined in Eq. (21).

$$\zeta\left(y, \hat{y}\right) = \sum_{i=1}^{2} y_i \log\left(\hat{y}_i\right) \tag{21}$$

Hyperparameters and implementation details

The key training settings and implementation choices are reported in Table 2. In short, the model was implemented using Tensorflow/Keras, with reproducibility guaranteed by setting the random seed to 0. The input MRI slices were of size $168 \times 168 \times q$ and trained for 50 epochs with a batch size of 8, where the training set was split into training (90%) and validation (10%) sets. Backbone: EfficientNetB1 (Pretrained on ImageNet) and multi-head self-attention (four heads, key dimension = 64) and residual bottleneck blocks, Cross-scale feature learning. The model was trained via Adam by using categorical cross-entropy as loss, and regularization was ensured by batch normalization and dropout. For data augmentation, which was more severe on the minority classes, e.g., AD, small rotations, flips, Gaussian noise addition and scaling of intensities were applied. To cope with class imbalance, the performance was presented using macro-averaged metrics.

Results and discussion

This paper applies the proposed method to the ADNI and ADNI1 MRI images to classify the images into four and three clinical categories: CN, MCI, MC and AD for ADNI, and CN, MC and AD for ADNI1. The datasets ADNI and ADNI1 were divided into 80% training, 10% validation and 10% testing. Although

Table 2. Hyperparameters and implementation details of the proposed model.

Parameter / Setting	Value / Description
Random Seed (SEED)	0 (applied to Python, NumPy, and TensorFlow for reproducibility)
Input Image Size	168 × 168 pixels
Number of Channels	3 (RGB slices from structural MRI)
Batch Size	8
Number of Epochs	50
Validation Split	0.10 (10% of the training set used for validation)
Number of Classes	4 (Cognitively Normal (CN), Mild Cognitive Impairment (MCI), Mild Cognitive (MC), Alzheimer's Disease (AD))
Backbone Network	EfficientNetB1 (pre-trained on ImageNet, fine-tuned on ADNI/ADNI1)
Attention Mechanism	Multi-Head Self-Attention (MHSA) with 4 heads, key dimension = 64
Residual Module	Bottleneck residual block with 1×1 and 3×3 convolutions + identity/projection paths
Activation Function	ReLU (nonlinear layers), SoftMax (final classification layer)
Loss Function	Categorical Cross-Entropy
Optimizer	Adam
Regularization	Dropout (0.1 in MHSA, 0.3 before the classification layer), Batch Normalization in CNNs
Data Augmentation	Random rotations (±10°), horizontal/vertical flips, Gaussian noise ($\sigma \leq 0.01$), intensity scaling (±10%) — applied more frequently to minority classes (esp. AD)
Evaluation Metrics	Accuracy, Precision, Recall, F1-score, Specificity (macro-averaged due to imbalance)

the ADNI1 scans were available in three anatomical planes (axial, coronal, and sagittal), the ADNI scans were included only in axial slice views. The ADNI1 dataset included 1113 of 476 AD and MC cases, and CN cases, while 896 of 64 AD and 2240 of 3200 MCI and MC cases are available in the ADNI dataset.

Training and validation accuracy and loss were recorded over 50 epochs for all models as measures of classification performance and generalization across anatomical planes. As the results show, the ADNI1 axial plane, Fig. 3(a) and the ADNI1 sagittal plane, Fig. 3(c), showed the best validation results, both reaching 98.46.

Their training and validation curves converged steadily with little variance, revealing good generalization. On the other hand, a similar pattern was observed in the ADNI1 coronal plane, Fig. 3(b), with a peak validation accuracy of 98.15. Finally, the training and validation of the ADNI dataset is shown in Fig. 3(d).

For further performance analysis and to interpret the class-wise predictions, confusion matrices were calculated for each model on the test set. These matrices allow one to gauge how well the models were able to differentiate between clinical categories. Fig. 4 illustrates the confusion matrices for each of the four models: ADNI1 axial, ADNI1 coronal, ADNI1 sagittal, and ADNI.

Furthermore, Table 3 illustrates the classification results for both the ADNI and ADNI1 datasets. Performance was assessed with macro-average performance metrics due to class imbalance and to give equal importance to categories. The macro-averaging of any performance metric (in this case, accuracy, precision, recall and F1-score) computes the metric by calculating it independently per class and then taking th e average of these values (without weights based on population). This strategy guarantees that each class, in particular minority ones like AD, gives no less contribution to the final score, even in the case of a big class imbalance. On the ADNI1 dataset, on axial slices, the model obtained a macro-averaged accuracy of 99.13, precision of 98.85, recall of 99.23, specificity of 99.61, and F1-score of 99.04. Coronal and sagittal slices also turned out to be strong, both having the same accuracy of 98.27. The coronal view had a precision, recall, F1 score, and specificity of 98.24, specificity of 99.05, whereas the sagittal view showed slightly lower values: precision of 98.07, recall of 97.84, specificity of 99.12, and F1-score of 97.95.

The confusion matrices on ADNI1 exhibit good per-class performance for all views. MCI was also labeled appropriately for 111 of 112 (misclassified as CN in only one) cases. CN was accurately predicted in 70 of 72 cases, with little confusion about AD. All 48 ADs were diagnosed accurately across both views, showing strong diagnostic sensitivity and specificity. These findings emphasize the broad applicability of the model across different anatomical orientations and its robustness to discriminate clinically significant categories.

On the ADNI dataset, the model maintained high performance even under a more imbalanced class distribution, especially when there are few AD cases. The general accuracy was 99.37%, while the macro-averaged precision and recall were 97.21 and 97.31, respectively. The specificity was even

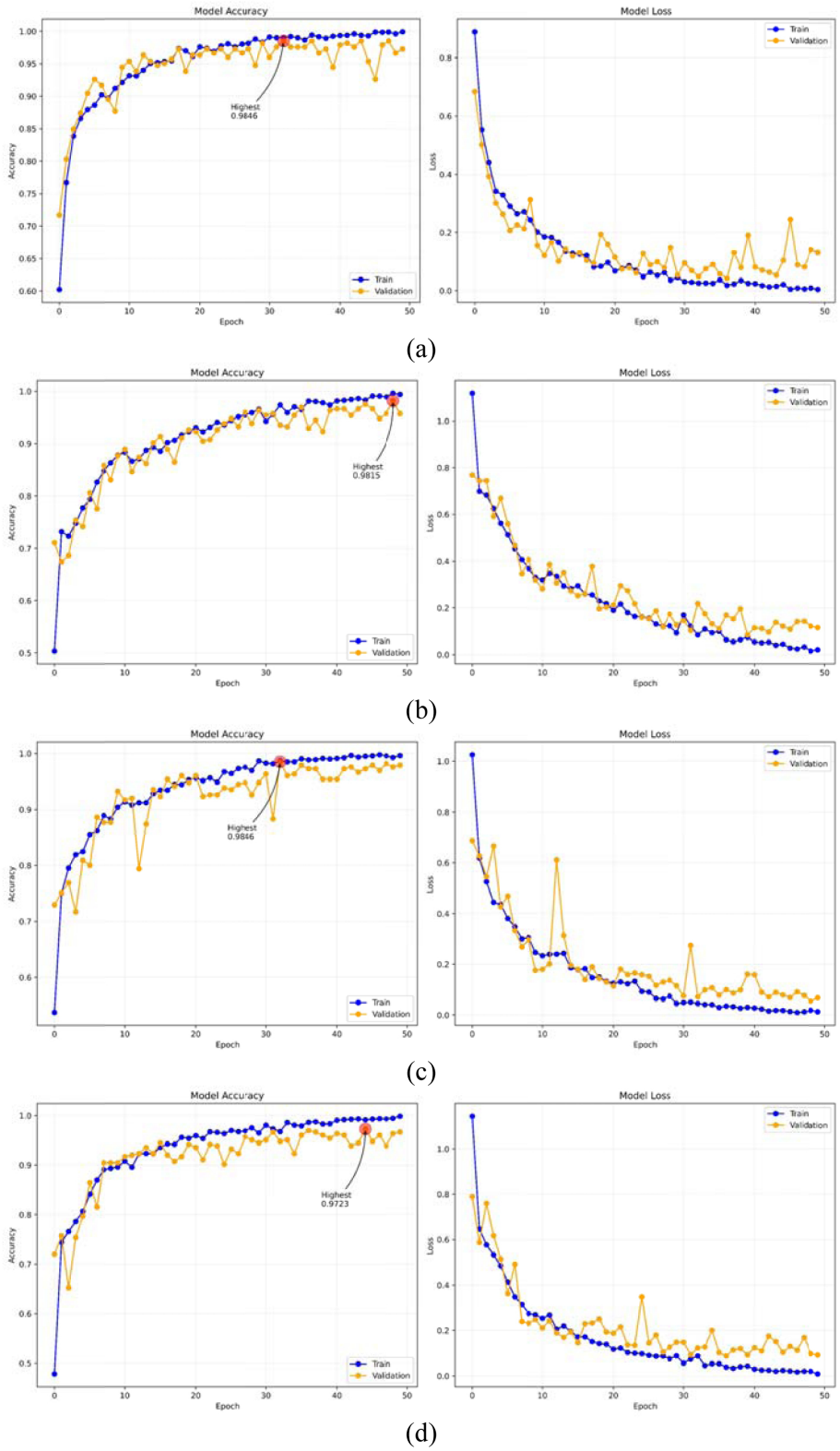

Fig. 3. Training and validation accuracy and loss: (a) ADNI1 sagittal plane, (b) ADNI1 coronal plane, (c) ADNI1 axial plane, and (d) ADNI dataset.

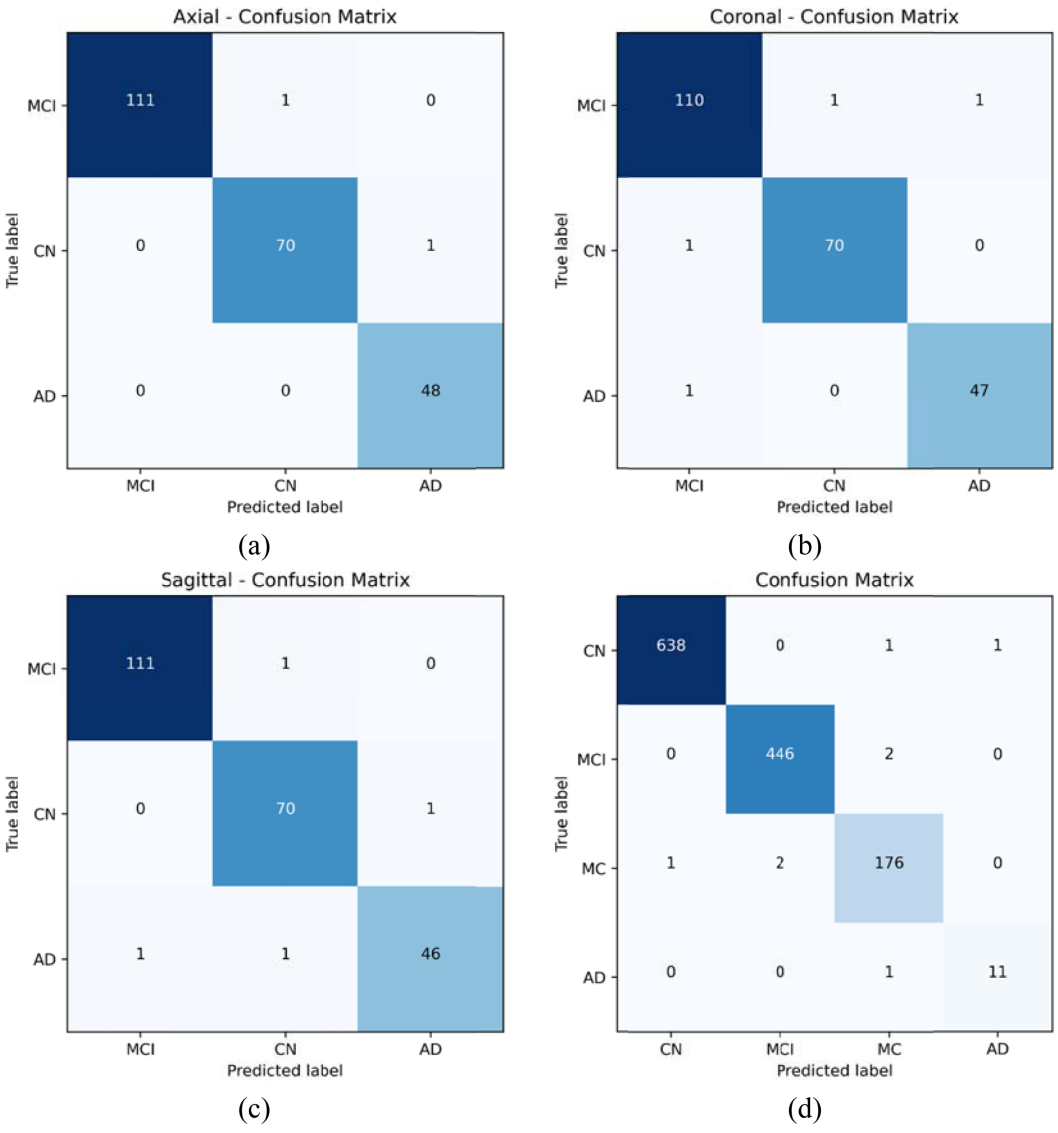

Fig. 4. Confusion matrices: (a) Axial, (b) Coronal, (c) Sagittal, and (d) ADNI.

Table 3. Classification performance on the ADNI and ADNI1 datasets' images.

Dataset	Slice	Accuracy	Precision	Recall	Specificity	F1 score
ADNI	Axial	99.37%	97.21%	97.31%	99.79%	97.26%
ADNI1	Axial	99.13%	98.85%	99.23%	99.61%	99.04%
	Coronal	98.27%	98.24%	98.24%	99.05%	98.24%
	Sagittal	98.27%	98.07%	97.84%	99.12%	97.95%

higher at 99.79, and the macro-averaged F1 score reached 97.26. The majority of the misclassified time points migrated between neighboring cognitive status groups (e.g., MCI to MC and CN to AD), which we attribute to the small morphological differences between these groups and the natural extent of disease progression.

These results confirm the robustness and high diagnostic performance of the model in different datasets and anatomical planes. The small preeminence of the coronal plane in the ADNI1 set might reflect the better visualization of the medial temporal structures that are frequently involved in the initial steps of AD. The good performance of the axial plane in the two datasets also indicates that this view is very informative for clinical classification in the presence of class imbalance.

Table 4 presents a comparison to other methods on multiple datasets and tasks. The reported metrics are accuracy (Acc), sensitivity (Sen), specificity (Spe),

Table 4. Classification performance on the ADNI and ADNI1 datasets.

Reference	Method	No. of Classes	Dataset	Acc (%)	Sen (%)	Spe (%)	Pre (%)	F1 (%)
Huang et al.[26]	2D CNN	2 classes	ADNI-1	96.48	97.83	93.98	–	–
			ADNI-2	96.3	93.71	97.99	–	–
Xin et al.[30]	MDBO-SVM	3 classes	ADNI	81.7	90.8	81.5	–	–
Rani and Baulkani[18]	ARSA-CA pre-trained SqueezeNet	4 classes	ADNI	98.3	98.1	98.7	98.9	98.6
		3 classes	OASIS	98.4	98.3	98.5	98.8	98.2
Garg et al.[29]	Statistical features + kNN	3 classes	ADNI	94.2	91.3	95.6	–	–
		3 classes	OASIS	94.9	92.2	96	–	–
Hassan et al.[32]	Dual-Stream	3 classes	ADNI1	98.8	98.8	98.8	–	–
Tekumudi and Ramanathan[25]	DenseNet121 + ViT	4 classes	ADNI	91.26	–	–	–	–
Proposed	Cross-Scale Feature	3 classes	ADNI1	99.13	99.61	99.23	98.85	99.04
		4 classes	ADNI	99.37	97.21	97.31	99.79	97.26

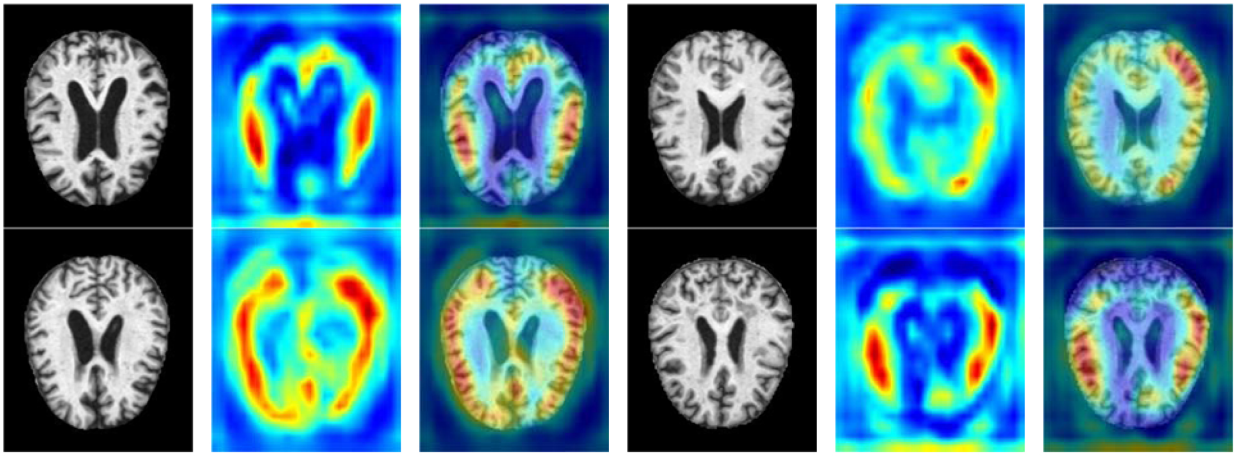

Fig. 5. Grad-CAM outputs with (a, d) original image; (b, e) Grad-CAM heatmaps; and (c, f) overlay images. Warmer colors indicate regions that contribute most to predictions computed by the model.

precision (Pre), and F1-score (F1) for three-class and four-class classification tasks in ADNI, ADNI1, and OASIS datasets. The proposed method was tested on a 3-class classification over the ADNI1 and a 4-class classification over the ADNI dataset.

For the three-class task on the ADNI1 dataset, the proposed method achieves Acc = 99.13%, Sen = 99.61% and Spe = 99.23%, which is better than all other methods, including the state-of-the-art Dual-Stream model of,[32] with a value of 98.8% for all measurements. Our proposed model yields substantial performance gains in terms of sensitivity and F1-score, two important indicators of clinical appropriateness, against traditional methods such as kNN[29] and MDBO-SVM,[30] with reported accuracies of 94.2% and 81.7%, respectively.

On the other hand, for the ADNI dataset, the proposed method achieved a superior Acc of 99.37%, Pre of 99.79%, and F1-score of 97.26% for the four-class classification. This is a substantial improvement over models, such as DenseNet121 + ViT,[25] with only 91.26% accuracy. For even the ARSA-CA SqueezeNet model,[18] which reported 98.3% accuracy, the proposed model performs better in all the available performance measures, especially in precision that pertains to the confidence of predictions in multi-class scenarios.

These findings also illustrate the effectiveness of the designed cross-scale feature extraction method. The stable, strong performance in both data and classification difficulty may indicate that the model not only interprets fine-scale anatomy but also generalizes well across cognitive levels. The wide gap in F1 scores implies balanced performance to minimize both false positives and false negatives, which is important for medical diagnostic responsibilities.

Finally, Grad-CAM was utilized on representative test MRI images to interpret the proposed model further. Grad-CAM heatmaps were created based on the final convolutional layer and applied to the respective original MRI slices to visualize spatial maps of the most effective regions for prediction. Fig. 5 shows the raw MRI images, their Grad-CAM generated heatmaps and overlay visualizations. The warmer

Table 5. Ablation study on the ADNI and ADNI1 datasets (mean ± standard deviation (SD)).

Method	No. of Classes	Dataset	Acc (%)	Sen (%)	Spe (%)	Pre (%)	F1 (%)
Backbone	3	ADNI1	96.42 ± 0.31	95.88 ± 0.47	96.71 ± 0.28	96.15 ± 0.36	96.01 ± 0.41
Cross-Scale (no attention)	3	ADNI1	97.85 ± 0.25	97.66 ± 0.33	97.93 ± 0.29	97.72 ± 0.26	97.69 ± 0.32
Attention (no cross-scale)	3	ADNI1	98.21 ± 0.22	98.04 ± 0.30	98.35 ± 0.26	98.11 ± 0.24	98.07 ± 0.27
Residual Bottleneck only	3	ADNI1	98.76 ± 0.19	98.59 ± 0.27	98.81 ± 0.23	98.64 ± 0.21	98.61 ± 0.25
Proposed (Cross-Scale, Attention, and Residual Bottleneck)	3	ADNI1	99.13 ± 0.14	99.61 ± 0.19	99.23 ± 0.17	98.85 ± 0.16	99.04 ± 0.18
Backbone only	4	ADNI	96.07 ± 0.36	94.38 ± 0.52	95.56 ± 0.31	96.91 ± 0.42	95.12 ± 0.46
Cross-Scale (no attention)	4	ADNI	97.63 ± 0.28	95.42 ± 0.38	96.12 ± 0.30	98.47 ± 0.33	96.89 ± 0.35
Attention (no cross-scale)	4	ADNI	98.11 ± 0.25	96.02 ± 0.34	96.73 ± 0.28	98.95 ± 0.29	97.47 ± 0.31
Residual Bottleneck only	4	ADNI	98.65 ± 0.21	96.78 ± 0.30	97.02 ± 0.25	99.23 ± 0.26	97.50 ± 0.28
Proposed (Cross-Scale, Attention, and Residual Bottleneck)	4	ADNI	99.37 ± 0.16	97.21 ± 0.23	97.31 ± 0.20	99.79 ± 0.18	97.26 ± 0.21

colored (red and yellow) regions represent the areas of higher relevance for the model's decision-making, while cooler colours (blue) are less contributing ones.

The Grad-CAM heatmaps show that the model attends more to the anatomical and clinically relevant regions of the brain. High-intensity regions appear to be concentrated in the cortical and periventricular areas, particularly around the lateral ventricles, as well as adjacent white matter. Such areas are typical regions for pathological alterations in neurological diseases, which would suggest that the model has learned discriminative and disease-relevant features.

The heatmaps display spatially related and symmetric patterns in much of the cortex of both hemispheres, indicating stable attention and not randomness or spurious activations. Interestingly, the overlaid Grad-CAM explanations exhibit a good match with brain tissue for the highlighted areas, and low response in background and non-brain regions.

This observation supports that the model's decisions are based on pertinent neuroanatomy, not spurious aspects in the image. In general, the Grad-CAM analysis contributes qualitative evidence of model explanability and enhances confidence and clinical interpretability of our framework.

Ablation study

To quantitatively analyze the impact of each architectural module, we performed a full ablation study over the ADNI1 (three classes, CN, MCI, and AD) and ADNI (four classes, CN, MCI, MC, and AD) datasets. The experiment dissects the contributions of the cross-scale feature extraction, multi-head attention and residual bottleneck module in a systematic manner by comparing the backbone network against its gradually strengthened versions. In addition to reporting the Mean performance metrics, we also report the standard deviations of these measurements over multiple runs to describe the robustness and reproducibility, in terms of model performance. This design has the advantage that performance improvements are not only apparent in point estimates but also statistically credible over trials. Table 5 illustrates the effectiveness of each architectural component, which is quantified in the ablation study for both the ADNI and ADNI1 datasets.

These values clearly indicate that every module is a source of demonstrable improvements. The backbone gives only a strong performance baseline; the sensitivities and F1-scores are pretty modest for each technique. The contribution of either the cross-scale feature extraction or the attention to sensitivity and specificity is upward 2% respectively, while the residual bottleneck could improve model stability and accuracy. Crucially, the uncertainty intervals are all narrow, thus illustrating that the performance gap is not random. These observations suggest that architectural design choices do indeed make a substantial difference to the performance of the model.

In the final aggregated rank method, it is observed that our proposed method attains a superior score among others in terms of accuracy, sensitivity and F1-score despite small standard deviations, which confirms its statistical stability. On ADNI1, the model yields near-ceiling performance, and on the more intricate dataset ADNI four-class, it achieves high precision and balanced generalization among classes. The ablation study, therefore, not only emphasizes the increased benefits from each module but also a new normalised benefit of combined features. These results highlight that the proposed framework is not only accurate but also robust, making it more appropriate for practical and clinical use.

Conclusions

In this paper, we proposed a Hybrid Deep Network for Structural MRI-based AD classification that

incorporated residual bottleneck modules, EfficientNetB1 as the backbone and cross-scale self-attention to model local anatomical patterns and long-range information jointly. The architecture overcomes shortcomings of the traditional CNNs, due to their inability to capture subtle global brain changes that are essential for early AD diagnosis. (i) We developed a new hybrid architecture of EfficientNetB1 with residual attention and cross-scale feature learning, which, to the best of our knowledge, is the first attempt to incorporate AD classification. (ii) The aggregated macro-averaged accuracy for the model is 99.13% (ADNI1) and 99.37% (ADNI), which outperforms several state-of-the-art methods on both three-class and four-class classification tasks. (iii) The model showed strong generalization performance across the anatomical planes — axial, sagittal and coronal — indicating the robustness of the model in capturing clinically meaningful features irrespective of view. (iv) The attention maps indicated biologically plausible areas (e.g., hippocampus, medial temporal lobe), facilitating interpretability and offering potential clinical utility in cross-validating automated decisions. (v) Ablation studies validated the contribution of each architectural module (i.e., cross-scale attention, residual bottleneck, and EfficientNet backbone). With all modules incorporated, the best stability and accuracy can be achieved. However, according to previous studies, these deep learning models usually suffer a 5-10% decrease in performance when transferring from well-designed datasets to real-world clinical settings. Accordingly, we expect that the true performance of our model may be in the 90-94% range, although acknowledging that our model has approximately 99% accuracy on ADNI and ADNI1. In future work, we will pursue multi-site external validation and domain adaptation approaches to establish robustness in light of clinical variability. Future work will seek to (i) validate the framework on multi-site, multi-protocol clinical datasets in order to guarantee robustness in real situations. (ii) We intend to validate our evaluation on external datasets (e.g., AIBL, OASIS, and other multicenter cohorts) and also confirm the robustness of our method. (iii) Generalization of the model to multimodal integration (sMRI + PET + genetics + clinical data analysis) for a more comprehensive evaluation of AD progression. (iv) Investigating more sophisticated imbalance-aware training techniques (e.g., class-weighted or focal-loss) to further improve the minority-class detection. (v) Conducting usability and workflow integration studies to assess the clinical relevance of the framework for neurologists and radiologists.

Authors' declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and images that are not ours have been included with the necessary permission for republication, which is attached to the manuscript.
- No animal studies are present in the manuscript.
- Author(s) signed on ethical consideration's approval.
- Ethical Clearance: The project was approved by the local ethical committee at Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia.

Authors' contributions statement

E. H. M. participated in the concept and design, model structure design, data preparation, and writing of the original draft. F. H. N. project oversight, overall study design, review and editing of the manuscript, model generation, data curation, formal analysis, final manuscript review, proofreading, and critical revision. S. M. K. contributed to optimizing deep learning, evaluating the models, data analysis, visualization and resources. M. H. K. helped with the literature review, manuscript structuring, and validation results. D. A. M. gave technical support and helped implement the evaluation metrics.

Data availability statement

The data supporting the findings of this study are completely available on request. ADNI and ADNI1 datasets: https://adni.loni.usc.edu/data-samples/adni-data.

References

1. Andrade-Guerrero J, Santiago-Balmaseda A, Jeronimo-Aguilar P, Vargas-Rodríguez I, Cadena-Suárez AR, Sánchez-Garibay C, *et al.* Alzheimer's disease: an updated overview of its genetics. Int J Mol Sci. 2023 Feb 13;24(4):3754. https://doi.org/10.3390/ijms24043754.
2. Tahami Monfared AA, Byrnes MJ, White LA, Zhang Q. Alzheimer's disease: epidemiology and clinical progression. Neurol. Ther. 2022 Jun;11(2):553–69. https://doi.org/10.1007/s40120-022-00338-8.
3. Alzheimer's Association. 2019 Alzheimer's disease facts and figures. Alzheimers Dement. 2019 Mar;15(3):321–87. https://doi.org/10.1016/j.jalz.2019.01.010.
4. Liss JL, Seleri Assunção S, Cummings J, Atri A, Geldmacher DS, Candela SF, *et al.* Practical recommendations for timely,

accurate diagnosis of symptomatic Alzheimer's disease (MCI and dementia) in primary care: a review and synthesis. J Intern Med. 2021 Aug;290(2):310–34. https://doi.org/10.1111/joim.13244.
5. Van Der Flier WM, de Vugt ME, Smets EM, Blom M, Teunissen CE. Towards a future where Alzheimer's disease pathology is stopped before the onset of dementia. Nat. Aging. 2023 May;3(5):494–505. https://doi.org/10.1038/s43587-023-00404-2.
6. Hechkel W, Helali A. Unveiling Alzheimer's Disease Early: A Comprehensive Review of Machine Learning and Imaging Techniques: W.Hechkel, A.Helali. Arch Comput Methods Eng. 2025 Jan;32(1):471–84. https://doi.org/10.1007/s11831-024-10179-3.
7. Dhabalia R, Kashikar SV, Parihar PS, Mishra GV, PARIHAR P. Unveiling the intricacies: a comprehensive review of magnetic resonance imaging (MRI) assessment of T2-weighted hyperintensities in the neuroimaging landscape. Cureus. 2024 Feb 24;16(2). https://doi.org/10.7759/cureus.54808.
8. Alsubaie MG, Luo S, Shaukat K. Alzheimer's disease detection using deep learning on neuroimaging: a systematic review. Mach Learn Knowl Extr. 2024 Feb 21;6(1):464–505. https://doi.org/10.3390/make6010024.
9. Najjar FH, Hassan NB, Kadum SA. Hybrid Deep Learning Model for Hippocampal Localization in Alzheimer's Diagnosis Using U-Net and VGG16. Int J Robot Control Syst. 2025 Apr 1;5(2). https://doi.org/10.31763/ijrcs.v5i2.1739.
10. El-Assy AM, Amer HM, Ibrahim HM, Mohamed MA. A novel CNN architecture for accurate early detection and classification of Alzheimer's disease using MRI data. Sci Rep. 2024 Feb 12;14(1):3463. https://doi.org/10.1038/s41598-024-53733-6.
11. Hassanin M, Anwar S, Radwan I, Khan FS, Mian A. Visual attention methods in deep learning: An in-depth survey. Inf Fusion. 2024 Aug 1;108:102417. https://doi.org/10.1016/j.inffus.2024.102417.
12. Hu Y, Deng L, Wu Y, Yao M, Li G. Advancing spiking neural networks toward deep residual learning. IEEE Trans. Neural Netw Learn Syst. 2024 Feb 8;36(2):2353–67. https://doi.org/10.1109/TNNLS.2024.3355393.
13. Kadhim KA, Mohamed F, Najjar FH, Salman GA. Early diagnose Alzheimer's disease by convolution neural network-based histogram features extracting and canny edge. Baghdad Sci J. 2024 Feb 25;21(2 (SI)):0643–0643. https://doi.org/10.21123/bsj.2024.9740.
14. Madhat IM, Kadhim KN, Mohamed F, Shafry Mohd Rahim M, Najjar FH, Ramadhan AJ. Diagnosing Alzheimer's disease severity: A comparative study of deep learning algorithms. InBIO Web of Conf., EDP Sciences. 2024 (Vol. 97, p. 00102). https://doi.org/10.1051/bioconf/20249700102.
15. Mao X, Chen Y, Jia X, Zhang R, Xue H, Li Z. Context-aware robust fine-tuning. Int J Comput Vis. 2024 May;132(5):1685–700. https://doi.org/10.1007/s11263-023-01951-2.
16. Kadhim KA, Mohamed F, Sakran AA, Adnan MM, Salman GA. Early diagnosis of Alzheimer's disease using convolutional neural network-based MRI. Mal. J. Fund. Appl. Sci. 2023 May 26;19(3):362–8. https://doi.org/10.11113/mjfas.v19n3.2908.
17. Abed MT, Fatema U, Nabil SA, Alam MA, Reza MT. Alzheimer's disease prediction using convolutional neural network models leveraging pre-existing architecture and transfer learning. In 2020 joint 9th international conference on informatics, electronics & vision (ICIEV) and 2020 4th international conference on imaging, vision & pattern recognition (icIVPR) 2020 Aug 26 (pp. 1–6). IEEE. https://doi.org/10.1109/ICIEVicIVPR48672.2020.9306649.
18. Rani KE, Baulkani S. Alzheimer disease classification using optimal clustering based pre-trained SqueezeNet model. Biomed. Signal Process. Control. 2025 Feb 1;100:107032. https://doi.org/10.1016/j.bspc.2024.107032.
19. Hassan N, Musa Miah AS, Shin J. Residual-based multi-stage deep learning framework for computer-aided Alzheimer's disease detection. J Imaging. 2024 Jun 11;10(6):141. https://doi.org/10.3390/jimaging10060141.
20. Kumar A, Pradeep S, Arora K, Sreeram G, Pankajam A, Patil T, Sahu A. Feature fusion based deep learning model for Alzheimer's neurological disorder classification. Neurosci Inform. 2025 Jun 1;5(2):100196. https://doi.org/10.1016/j.neuri.2025.100196.
21. Tay Y, Dehghani M, Bahri D, Metzler D. Efficient transformers: A survey. ACM Comput Surv. 2022 Dec 7;55(6):1–28. https://doi.org/10.1145/3530811.
22. Alp S, Akan T, Bhuiyan MS, Disbrow EA, Conrad SA, Vanchiere JA, Kevil CG, Bhuiyan MA. Joint transformer architecture in brain 3D MRI classification: its application in Alzheimer's disease classification. Sci Rep. 2024 Apr 18;14(1):8996. https://doi.org/10.1038/s41598-024-59578-3.
23. Akan T, Alp S, Bhuiyan MA. Vision transformers and Bi-LSTM for Alzheimer's disease diagnosis from 3D MRI. In2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE) 2023 Jul 24 (pp. 530–535). IEEE. https://doi.org/10.1109/CSCE60160.2023.00093.
24. Lozupone G, Bria A, Fontanella F, Meijer FJ, De Stefano C. AXIAL: Attention-based eXplainability for Interpretable Alzheimer's Localized Diagnosis using 2D CNNs on 3D MRI brain scans. arXiv preprint arXiv:2407.02418. 2024 Jul 2. https://doi.org/10.48550/arXiv.2407.02418.
25. Tekumudi VS, Ramanathan A. Hybridvit: An approach for alzheimer's disease classification with adni neuroimaging data. SN Comput. Sci. 2025 Apr 6;6(4):356. https://doi.org/10.1007/s42979-025-03862-0.
26. Huang H, Pedrycz W, Hirota K, Yan F. A multiview-slice feature fusion network for early diagnosis of Alzheimer's disease with structural MRI images. Inf Fusion. 2025 Jul 1;119:103010. https://doi.org/10.1016/j.inffus.2025.103010.
27. Abdelaziz M, Wang T, Anwaar W, Elazab A. Multi-scale multimodal deep learning framework for Alzheimer's disease diagnosis. Comput Biol Med. 2025 Jan 1;184:109438. https://doi.org/10.1016/j.compbiomed.2024.109438.
28. Li S, Lin S, Tu Y, Qiao J, Xiao S. Unified multi-protocol MRI for Alzheimer's disease diagnosis: Dual-decoder adversarial autoencoder and ensemble residual shrinkage attention network. Biomed. Signal Process. Control. 2025 Jul 1;105:107660. https://doi.org/10.1016/j.bspc.2025.107660.
29. Garg N, Choudhry MS, Bodade RM, Alzheimer's Disease Neuroimaging Initiative. Alzheimer's disease detection through wavelet-based shifted elliptical local binary pattern. Biomed. Signal Process Control. 2025 Feb 1;100:107067. https://doi.org/10.1016/j.bspc.2024.107067.
30. Xin Y, Sheng J, Zhang Q, Song Y, Wang L, Yang Z. A novel diagnosis method utilizing MDBO-SVM and imaging genetics for Alzheimer's disease. Comput Med Imaging Graph. 2025 Jul 1;123:102542. https://doi.org/10.1016/j.compmedimag.2025.102542.
31. Alinsaif S. DCA-Enhanced Alzheimer's detection with shearlet and deep learning integration. Comput Biol. Med. 2025 Feb 1;185:109538. https://doi.org/10.1016/j.compbiomed.2024.109538.

32. Hassan N, Miah AS, Suzuki T, Shin J. Gradual variation-based dual-stream deep learning for spatial feature enhancement with dimensionality reduction in early Alzheimer's disease detection. IEEE Access. 2025 Feb 14. https://doi.org/10.1109/ACCESS.2025.3542458.
33. Tuncer T, Dogan S, Subasi A. FiboNeXt: investigations for Alzheimer's disease detection using MRI. Biomed. Signal Process. Control. 2025 May 1;103:107422. https://doi.org/10.1016/j.bspc.2024.107422.
34. Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, *et al.* Alzheimer's disease Neuroimaging Initiative (ADNI) clinical characterization. Neurol. 2010 Jan 19;74(3):201–9. https://doi.org/10.1212/WNL.0b013e3181cb3e25.

تعلم الميزات عبر المقاييس لاكتشاف مرض الزهايمر باستخدام شبكات عميقة هجينة

عصام حيدر مجيد [1]، فلاح حسن نعمة [1، 2]، صفاء محسن كريم [1]، مناف حمزة كريم [1]، ضُحى عامر مهدي [1]

[1] قسم تقنيات شبكات وبرامجيات الحاسوب، المعهد التقني النجف، جامعة الفرات الأوسط التقنية، النجف، العراق.
[2] قسم الحوسبة الناشئة، كلية الحوسبة، الجامعة التكنولوجية الماليزية، جوهور باهرو، ماليزيا.

الخلاصة

مرض الزهايمر (AD) هو اضطراب عصبي تنكسي مزمن يُعد السبب الرئيسي للخرف. يُعتبر التشخيص المبكر أمرًا بالغ الأهمية لوقف تدهور القدرات المعرفية وتحسين النتائج السريرية، على الرغم من أن وسائل التشخيص التقليدية قد لا تكون قادرة على اكتشاف الأشكال العرضية المبكرة للمرض. في هذه الدراسة، نقترح نموذج تعلم عميق هجين يعتمد على(Residual Attention) ويستخدم (EfficientNetB1) لتحسين تصنيف مرض الزهايمر اعتماداً على بيانات التصوير بالرنين المغناطيسي الهيكلي المأخوذة من قواعد بيانات ADNI وADNI1. يُدمج هيكل الشبكة بين استخراج الميزات ثنائية الأبعاد (2D Convolutional Feature Extraction)، وآلية (Multi-head Self-Attention) لبناء معلومات السياق العالمية، بالإضافة إلى (Residual Learning)، مما يتيح للنموذج اكتساب الخصائص التشريحية المحلية وكذلك العلاقات الهيكلية العالمية للدماغ بشكل فعّال. وقد تفوق النموذج المقترح على العديد من النماذج المتقدمة من حيث دقة التشخيص، حيث وصل إلى دقة متوسطية كلية بنسبة 99.13% و99.37% على بيانات ADNI وADNI1 على التوالي، وذلك باستخدام ثلاث مستويات تشريحية (المستوى المحوري، والسهمي، والإكليلي). وقد أدى ذلك إلى تحديد مناطق بيولوجية قابلة للتفسير من خلال خرائط الانتباه، مما عزز من قابلية تفسير النموذج. تشير هذه النتائج إلى أن نموذجنا يُعد وسيلة فعالة وقابلة للتفسير لتصنيف مرض الزهايمر بدقة في مراحله المبكرة، مع إمكانية تطبيقه في البيئات السريرية. بالإضافة إلى ذلك، سيتم تقييم النموذج من حيث قابليته للاستخدام وفعاليته العملية في الاستخدام السريري لضمان أن يكون للنموذج قابلية كافية للتطبيق من قبل أطباء الأعصاب وأطباء الأشعة.

الكلمات المفتاحية: مرض الزهايمر، السمات عبر المقاييس، التصنيف، التعلم العميق، نموذج هجينة.