Cerevisia 50 (2026)

Contents lists available at SciVerse ScienceDirect

Cerevisia

EDA-Net: An Attention-Enhanced Convolutional Neural Network for Multi-Stage Alzheimer's Disease Diagnosis from MRI Images

Zimeng Yi [a,*]

[a] School of Life Sciences, Qingdao University, Qingdao 266071, China

Article info

Abstract

The accomplishment of recognizing Alzheimer's disease AD in its beginning stages especially identifying mild cognitive impairment MCI will cause the onset of the disease to occur later. This is very tough to perform due to the finesses associated with making early pathological changes visually undifferentiated. MRI imaging provides helpful information about the structural problems associated with AD, yet many current computer-aided diagnostic systems do not have fine enough representation of features or precise spatial location information. Therefore In response to these difficulties the authors propose EDA-Net Enhanced Deep Attention Network a newly developed Deep convolutional Neural Network to detect multiple stages of AD from MR images. In particular we are proposing an innovative component of EDA-Net called the Deep Attention DA Block which utilizes our newly developed Efficient Coordinate-Spatial Attention Mechanism ECSAM that incorporates Efficient Channel Attention ECA and Coordinate Attention CA and provides a mechanism for sequentially connecting the various channels of the DA Block with respect to collecting the inter-channel information as well as modeling the long range spatial dependencies within the network to focus dynamically on the areas of greatest discrimination in the areas of the MRI images. The DA Block uses SiLU activation and Group Normalization GN to provide both additional nonlinear representational ability and stability when training the network with small-batches of training data. EDA-Net has been validated through five multiple folds of cross-validation on a publicly available database containing 6400 MR images obtained from four consecutive clinical stages. Based on experimental results, EDA-Net is able to achieve a mean classification accuracy of 98.30%, with all other important metrics having an accuracy of over 98.38%, and a Kappa coefficient of 97.73%. This essentially makes EDA-Net superior in performance to traditional models, such as ResNet50 and DenseNet121. The results of the ablation studies have shown a synergy between ECA and CA components, and the confusion matrix has confirmed that this network has a very high level of class discrimination. Overall, EDA-Net has a unique architecture combined with a hybrid mechanism of attention to allow it to balance between high accuracy, efficient use of resources, and optimum performance. For all of these reasons, EDA-Net is a very robust deep learning solution to screen for the early stages of Alzheimer's disease. Future research will continue validating the generalizability of the model using more diverse clinical datasets and will be investigating ways to integrate different types of information sources to enable clinical translation.

Keywords: Alzheimer's disease; convolutional neural network; deep learning; attention mechanism

1. Introduction

Alzheimer's disease (AD), one of the most common forms of neurodegeneration worldwide, continues to present serious and increasing public health concerns because of the ever-increasing rates of incidence of this disease over time [1]. The natural process of Alzheimer's disease progresses in a gradual, steady manner from cognitive normality through subjective cognitive decline and through mild cognitive impairment (MCI) and then develops into AD dementia [2,3]. A large amount of evidence demonstrates how an effective clinical response in the early stages of AD, particularly when diagnosed during the MCI stage, has the potential to delay or leave some portion of the cognitive declines that occur during this disease, which makes these early phases a major point of opportunity to improve the long-term outcomes for patients with AD [4]. Therefore, the continued improvement of the diagnostic methods of Alzheimer's Disease to correctly differentiate between normal cognition, very mild impairment, mild impairment, and moderate impairment are extremely critical to level the playing field for developing individualized treatment approaches, recruiting patients to participate in clinical studies, and following how patients progress through the disease [5]. In general, at these very early phases of the disease, changes occurring pathologically within the brain are typically found to be difficult to identify because such changes are typically so subtle, scattered, and non-typical that the ability to complete a reliable clinical assessment and/or make a radiological interpretation of brain images is not achievable.

MRI is a type of medical imaging that uses a magnetic field and radio waves to produce detailed images of the inside of your body. MRI is considered to be the gold standard of imaging studies for Alzheimer's disease (AD) because it is non-invasive and does not expose patients to ionizing radiation; it also provides better contrast for soft-tissue structures than any other type of imaging technique, and it can produce multiple types of images from one MRI scan. Using structural MRI, researchers can determine how much each area of the brain shrinks (atrophies) over time. Regions that show significant atrophy early on (e.g., the hippocampus and entorhinal cortex) serve as sensitive biomarkers for AD [6, 7]. MRI provides many advantages over other imaging methods, such as PET scans, as it can be more easily obtained, is less expensive, is more reproducible, and therefore, is more suited to being used for population screening and longitudinal follow-up studies [8], which makes mining the pathologic data contained in MRIs and developing automated and objective analytical tools for computer-aided diagnosis of AD a critical area of research currently.

The rapid progress in deep learning, especially in convolutional neural networks, has brought about transformative advancements in intelligent medical image analysis [9–11]. Unlike traditional machine learning methods, which require extensive human intervention in the design of features, CNNs can automatically learn multi-layered hierarchical representations of features from the raw input data, allowing them to better understand complex and non-linear present-day illnesses [12]. There are numerous studies that show superior performance of deep learning models for binary classification tasks like distinguishing AD patients from healthy controls; however, it remains to be seen whether or not, these same methods will be successful for multi-stage AD classification tasks such as differentiating healthy elderly patients from those with different levels of AD severity [13–17]. In the process of creating and evaluating deep-learning models, there are three major limitations associated with these tasks: (1) repetitive downsampling operations can reduce the amount of fine texture features needed for classifying patients with AD; (2) standard 2D convolutional kernels are limited in their ability to model features that represent spatial dependencies over distance and (3) deep learning model interpretability, making them unequally suitable for clinical practice.

To systematically overcome these limitations and facilitate the advancement of accurate and timely identification of early stage AD, this study proposes a novel, deep-learning solution for AD diagnosis by summarizing the main contributions of this study:

1. The proposed Efficient Coordinate-Spatial Attention Module (ECSAM) efficiently calibrates information interactions between channels and captures long-range dependencies along key spatial coordinates to allow the model to focus on regions of the image most important to diagnosing Alzheimer's disease (AD). ECSAM combines Efficient Channel Attention (ECA) [18] and Coordinate Attention (CA) [19] into a hybrid sequentially organized attention module.
2. The Deep Attention Block (DA Block) is a new fundamental network block that builds on ECSAM. The DA Block incorporates the attention mechanism described above, as well as, the SiLU Activation and Group Normalization (GN). The DA Block was designed with the intent to increase the model's ability to learn non-linear representations and to provide greater training stability and generalization performance in small batch scenarios.
3. Using the DA Block as its core building component, we have created EDA-Net (Efficient Dual-Attention Network), a brand new model architecture designed from scratch and tailored to capture complex and subtle imaging biomarkers through the multi-stage progression of Alzheimer's disease.
4. We performed extensive experiments on a large public MRI data set through rigorous five-fold cross validation, comprehensive ablation studies, and benchmark comparisons to multiple state-of-the-art architectures including ResNet50 and ResNeXt50. EDA-Net outperformed all referenced models with the highest level of predictive accuracy and greatest robustness for four-class AD classification. Thus, EDA-Net offers computer-aided diagnostic support tools for the early identification of patients with Alzheimer's Disease.

2. Method

2.1 Overall Network Architecture

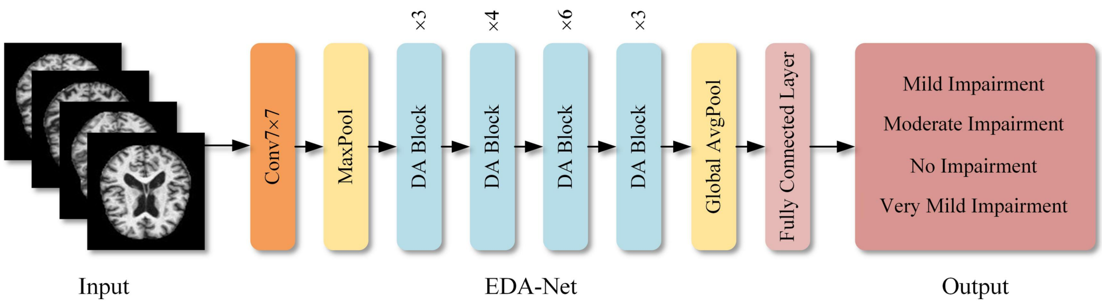

Figure 1: Overall Framework Diagram of the Model

The proposed end-to-end system for classifying multi-stage Alzheimer's disease (AD) based on MRI scans is described in Figure 1. In this pipeline, raw brain MRI scans are preprocessed, then features extracted and enhanced before producing four classifications of disease stage.

This proposed system will autonomously and intelligently map the MRI image data to the corresponding clinical stages of AD. The process begins with untrained MRI scans as the initial input, after which the scans will undergo several operations designed to standardize and increase the generalizability of the trained model. The first of these operations will be to standardize the resolution of all MRI images to 224×224 pixels. This standardization will be accomplished with the use of random horizontal flips (for augmentation of the dataset) prior to converting the images to tensor format suitable for deep learning models, and normalizing to assist with model convergence and stabilizing model training.

Once the MRI scans are preprocessed, they can be submitted to the core feature extraction network (EDA-Net) based on Deep Learning procedures. EDA-Net is an adaptation of the ResNeXt50 architecture [20], which was built around a concept called "cardinality" and grouped convolution methods to allow for networks that can be many times wider than a ResNet architecture while preserving the same number of parameters. The ResNeXt50 architecture reduced the effects of vanishing gradients in deep neural networks through stacking residual layers, producing the optimal architecture for the design of very high-performance feature extractors [20]. EDA-Net builds on the ResNeXt50 design with the addition of several new structures and improvements via its integration of Deep Attention Blocks (DA Blocks) into EDA-Net's backbone architecture.

The DA Block contains two significant features. First, it uses the Enhanced Channel-Spatial Attention Module (ECSAM), which consists of the architecture showed in Fig. 2. The ECSAM sequentially combines Efficient Channel Attention (ECA) and Coordinate Attention (CA). The ECA is effective at capturing the inter-channel dependency with one-dimensional (1D) convolution, while CA uses coordinate information to model the long-range spatial dependency via decomposed coordinate attention. For CA, the encoding and fusion of the attention is performed separately along the height and width directions. Using both of these mechanisms together allows the network to dynamically recalibrate the feature channels and attend to spatial regions closely related to AD pathology.

Second, the SiLU (Sigmoid-weighted Linear Units) activation function has been adopted in the DA Block in lieu of ReLU (Rectified Linear Units) to enable a more linear and smoother nonlinear behavior and an enhanced gradient flow. In addition, Group Normalization (GN) has been used in place of Batch Normalization (BN) to improve training stability and performance with small batch sizes, which is essential in medical imaging tasks since most datasets do not have a sufficient volume of data to support a large batch size.

The EDA-Net architecture combines and integrates DA blocks with enhanced features using attention mechanisms and well-designed network components to extract high level semantic understanding of the images of the Early Stage Alzheimer's Disease patient. The functionally superior component of the EDA-Net is is an end to output vector which is created by utilizing the pool of high level semantics. The categorical labels assigned are; No Impairment, Very Mild, Mild and Moderate Impairment, representing categories of AD severity. By designing the EDA-Net with the intended capture of early-stage Alzheimer's Disease related imaging biomarkers, the EDA-Net is refined through DA Block processing so that it optimally extracts imaging biomarker features and provides a classification system for multi-stage AD with a high degree of classification accuracy.

2.2 Enhanced Channel–Spatial Attention Module (ECSAM)

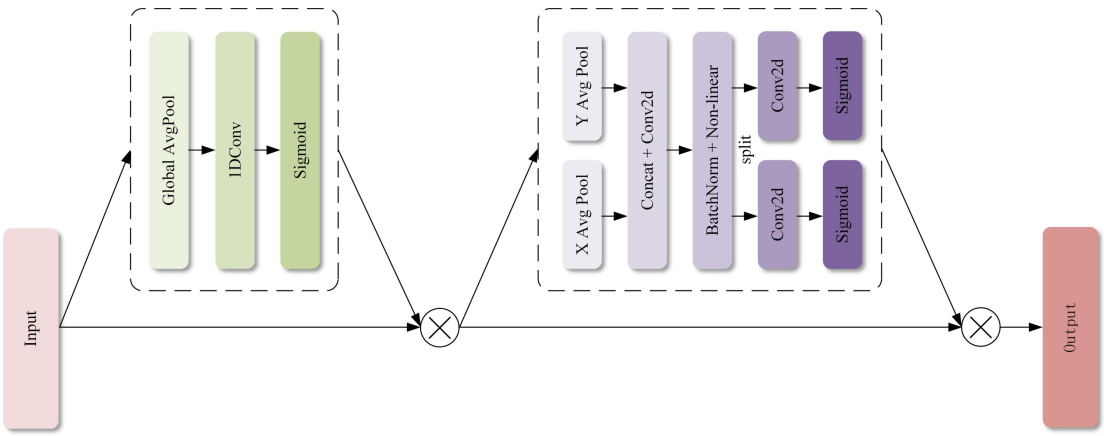

Figure 2: Enhanced Channel-Spatial Attention Module Architecture Diagram

ECSAM represents the main processing element of the DA Block (shown in Fig. 2), adding flexibility to the computation by using a series of processing steps that cooperatively optimize both channel-based and spatial-based attention using minimal additional processing complexity (i.e. computational cost).

The approach used in ECSAM is based upon two widely used types of attention mechanisms, namely Coordinate Attention (CA) and Efficient Channel Attention (ECA). By using these approaches in a cascading manner, ECSAM refines both channel-level and spatial representation by progressively improving them through direct channel and spatial modelling with very low computational cost and, therefore allowing for greater selectivity on discriminative areas of the image.

ECSAM first receives the input feature map F via the ECA module. The ECA module takes a lightweight approach and performs the attention mechanism using a 1D convolution with an adaptively sized kernel (i.e. k) that allows for the modelling of relationships between neighbouring channels throughout the entire channel dimension without causing any losses in information due to the reduction in dimensionality. Therefore, the processing step can be represented mathematically as follows:

$$F_c = M_c(F) \times F \tag{1}$$

where $\mathrm{M_c(F) = \sigma(IDConv(GAP(F)))}$ denotes the generated channel attention map, σ is the sigmoid activation function, GAP represents global average pooling, 1DConv denotes one-dimensional convolution, and $\times$ indicates channel-wise multiplication. This step effectively recalibrates channel importance and emphasizes more informative feature channels. Subsequently, the channel-enhanced feature F_c is fed into the Coordinate Attention (CA) submodule. Coordinate Attention (CA) distinguishes itself from other methods by embedding coordinate information where spatial position is decomposed into two independent components (vertical & horizontal) which preserve the long-distance relationship of spatial dependency. This is done in two steps:.

First Step: Coordinate Information Embedding - Split pooling the input feature into two separate feature maps based on the X and Y directions of the coordinate system.
Second Step: Using the concatenated feature maps, create attention maps for the X and Y directions, and multiply those attention maps with the original feature map.

The mathematical formulation is:

$$F_c^h = f^h(F_c) \times F_c \tag{2}$$

$$F_{out} = f^w(F_c^h) \times F_c^h \tag{3}$$

where f^h and f^w denote the coordinate attention mapping functions along the height (vertical) and width (horizontal) directions, respectively. Using this approach, the model can accurately identify geographical locations within the image that represent areas of interest for the specified target, including but not limited to areas of pathology. Therefore, the entire calculation for this module can be summarized as follows:

$$F_{out} = M_s^{ca}(M_c^{eca}(F) \times F) \times (M_c^{eca}(F) \times F) \tag{4}$$

where M_c^{eca} is the channel attention map generated by the ECA submodule, M_s^{ca} is the spatial attention map generated by the CA submodule, and $\times$ is element-wise multiplication.

The proposed two-step approach consists of an initial step of selecting specific feature channels and a second step of aligning those channels spatially. This combination of ECA and CA creates an interplay of complementary modulation between the two layers, whereby ECA yields a set of high-quality, highly distinctive feature channels for the next stage of fine classification and spatial localization. In addition, CA allows for increased emphasis to be placed on specific spatial areas represented by the selected channels. As a result of this two-pronged approach to adapting to the complexities of the Alzheimer's Disease pathology, ECSAM allows for an overall increase in the precision of classification by providing clear, interpretable representations of features that describe the disease's subtle characteristics.

2.3 Dual-Attention Residual Block (DA Block)

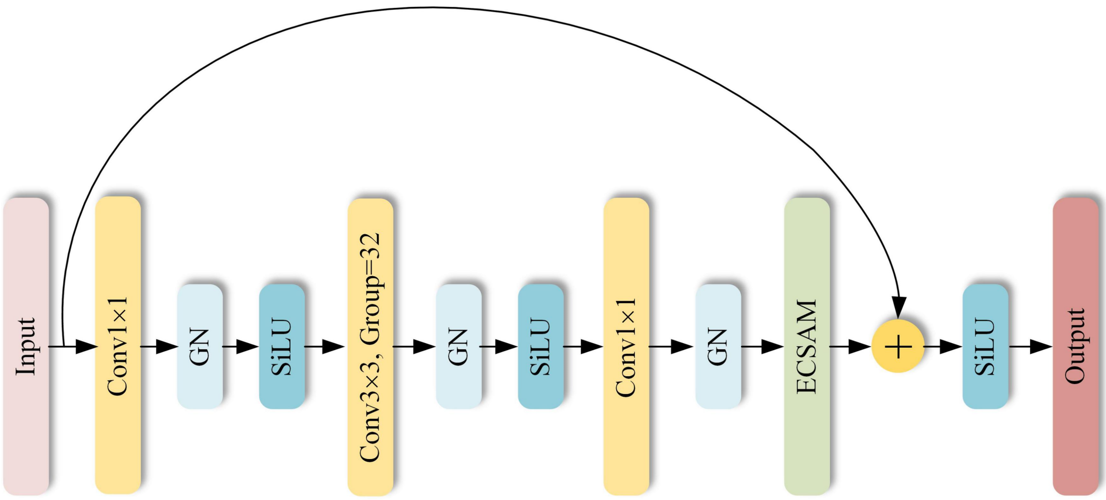

Figure 3: Framework Diagram of the Dual Attention Residual Structure Block

We will build a dual-attention residual block (DA Block). This dual-attention design has been shown in Fig 3. The DA Block uses Group Normalization [21] and SiLU [22] instead of using standard Batch Normalization (BN) and Rectified Linear Units (ReLU). Because Group Normalization is not dependent on batch size and provides more consistent results especially when training via small batch sizes compared to standard BN, it is considered a better choice. The SiLU nonlinearity will provide a smoother alternative to the ReLU activation function. It also helps to address the issue of vanishing gradients due to its ability to provide a non-zero gradient around the zero input value. The convolutional aspect of the DA Block will implement grouped convolutions (Group = 32), based on the ResNeXt-50 model, which can broaden the network's ability to represent data without a proportional increase in parameters. Each grouped convolution will be followed with group normalization and SiLU for a consistent normalization and activation style. In addition to the previous features, we will add an additional attention module (ECSAM) to the output of the main branch of the DA block prior to connecting to the residual connection in order to further focus and enhance features through recalibration and integration. This additional component will aid the model in effectively capturing the important content in the image as well as improving feature quality and discriminability. As a whole, this new block will increase representation and generalization ability of the models while maintaining their computational efficiency.

3. Experiments

3.1 Dataset Composition and Class Distribution

The MRI dataset for this research was taken from an established and well-known benchmark dataset that is publicly available for researchers interested in analysing brain scans through the use of Alzheimer's disease (AD) diagnostic imaging [8]. This benchmark dataset was created to help to overcome the common challenges of having very few data points available and having a class imbalance between the non-demented population and the advanced stages of AD, and has systematically encompassed all four major transformational phases (based on clinical research) of the development of AD; 1. No impairment 2. Very mild impairment 3. Mild impairment 4. Moderate impairment (Examples of each of the four clinical stages are presented in Figure 4). The training dataset for each of the four clinical categories contained at least 2,560 axial MRI images for sufficient and stable model training, and each image was generated using high-quality synthetic image generation methods. When examining how closely the synthetic images resemble real MRI images in terms of quantitative accuracy (lower FID scores, greater SSIM and PSNR scores and near ideal image sharpness), the quality of the TPG synthetic images exhibited a high level of consistency compared to real MRI images [23]. The combination of all this information creates a foundation of reliable, balanced and high quality imaging data for model training and development.

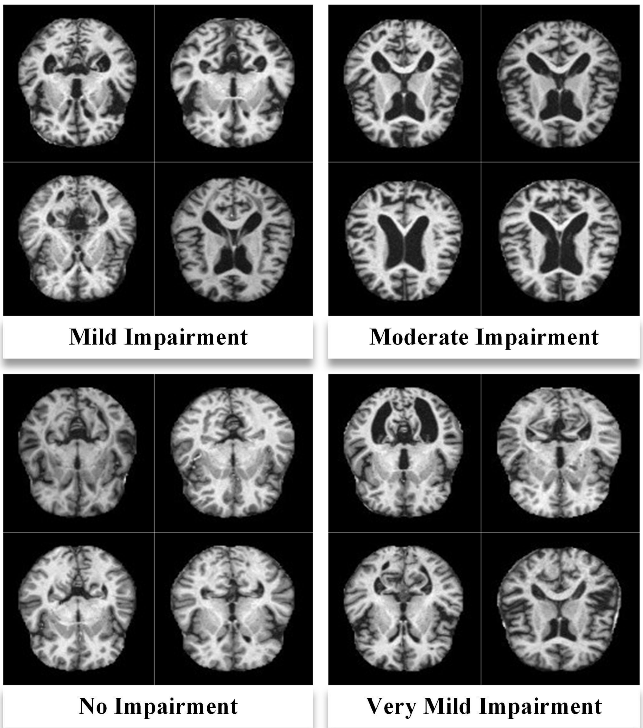

Figure 4: Illustrative Diagram of Four Categories of Data

3.2 Experimental Setup and Parameter Settings

The workstation used in all experimentation included an NVIDIA GeForce RTX 4090 GPU with 24 GB of memory and used PyTorch 1.12.1 as the deep learning framework. The model's performance was evaluated using five-fold stratified cross-validation, meaning that the dataset was divided into five equal parts, with four of these parts being used to train the model, while one part was utilised as a validation set in each fold. The final results from the five-fold cross-validation are reported as the average of the model's performance along with one standard deviation of that average to create a level of reliability in the results produced.

The training hyperparameters were configured as follows: the batch size was set to 32, the initial learning rate was 0.001, and the Adam optimizer was used with an exponential learning rate scheduler, where the decay factor γ was set to 0.96. Each model was trained for 160 epochs, and cross-entropy loss was adopted as the objective function. In the data preprocessing phase, horizontal flip copy 50% during horizontal flipping, random cropping into a 224×224 pixel image and the data were normalized to standard values. During the data testing phase, center cropping method was applied followed by the same method of the above normalisation method. Data preprocessing methods included random cropping into a 224 x 224 pixel image with normalisation, and centre cropping with normalisation was used for testing. The backbone of the model used was ResNeXt50 with attention-enhanced blocks; the activation function of ReLU was replaced with SiLU and in addition Batch Normalisation was replaced with Group Normalisation to ensure a controlled experiment. All of these methods were designed to provide the ability to compare experimental data.

3.3 Ablation Study

This paper presents a new type of Efficient Channel-Spatial Attention Module (ECSAM) as well as its application in EDA-Net which is an improved version of deep feature extraction network. The study includes a thorough analysis of the two core components of ECSAM, i.e., Efficient Channel Attention (ECA) and Coordinate Attention (CA). The results of the experiments conducted to analyze their combined effects on model performance are provided in Table 1. As can be seen from Table 1, even without any form of attention mechanism, the EDA-Net model produces a respectable accuracy rate of

95.92% when it makes predictions about patients who have been diagnosed with Alzheimer's Disease (AD). Adding CA to the model substantially improves accuracy to 97.93%. The reason why CA provides such a benefit is that it can effectively capture long-range spatial dependencies and therefore help to improve the ability of the model to accurately locate anatomically significant areas that are changing due to AD. In addition to including CA in the model's architecture, adding ECA was found to increase accuracy to 96.90%. This improvement can be attributed to ECA'S lightweight 1D convolutional operation, which can efficiently model the interactions between different channels of features when computing the optimal weight assignment for each channel's feature representation. By cascading ECA and CA to form the complete ECSAM module, the EDA-Net produced a superior model across all evaluation metrics, achieving an overall accuracy rate of 98.30% and Kappa value of 97.73%, and also exhibiting lower standard deviations than the previous versions, indicating improved stability. Therefore, these results demonstrate that ECA and CA do complement one another in the sense that ECA produces global refinement of the channel feature representations while CA focuses on improving the spatial localization of the anatomical features related to AD. The network's ability to recalibrate features across different dimensions and to more precisely identify the subtle and complex imaging biomarkers associated with early Alzheimer's disease (AD) (such as hippocampal atrophy and cortical thinning) is a major impact of this deep integration of AD social networks. As such, this deep integration provides a technical basis to support the development of a high-precision, multi-stage AD classification.

Table 1: Comparison of Ablation Experiment Results for the Efficient Channel-Spatial Attention Module (ECSAM)

ECA	CA	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)	Kappa (%)
×	×	95.92±0.34	96.10±0.32	96.10±0.34	96.09±0.33	94.55±0.46
×	√	97.93±0.41	98.02±0.40	98.02±0.40	98.02±0.40	97.23±0.54
√	×	96.90±0.46	97.04±0.43	97.05±0.44	97.04±0.44	95.86±0.61
√	√	98.30±0.12	98.37±0.13	98.39±0.11	98.38±0.12	97.73±0.16

3.4 Confusion Matrix Analysis

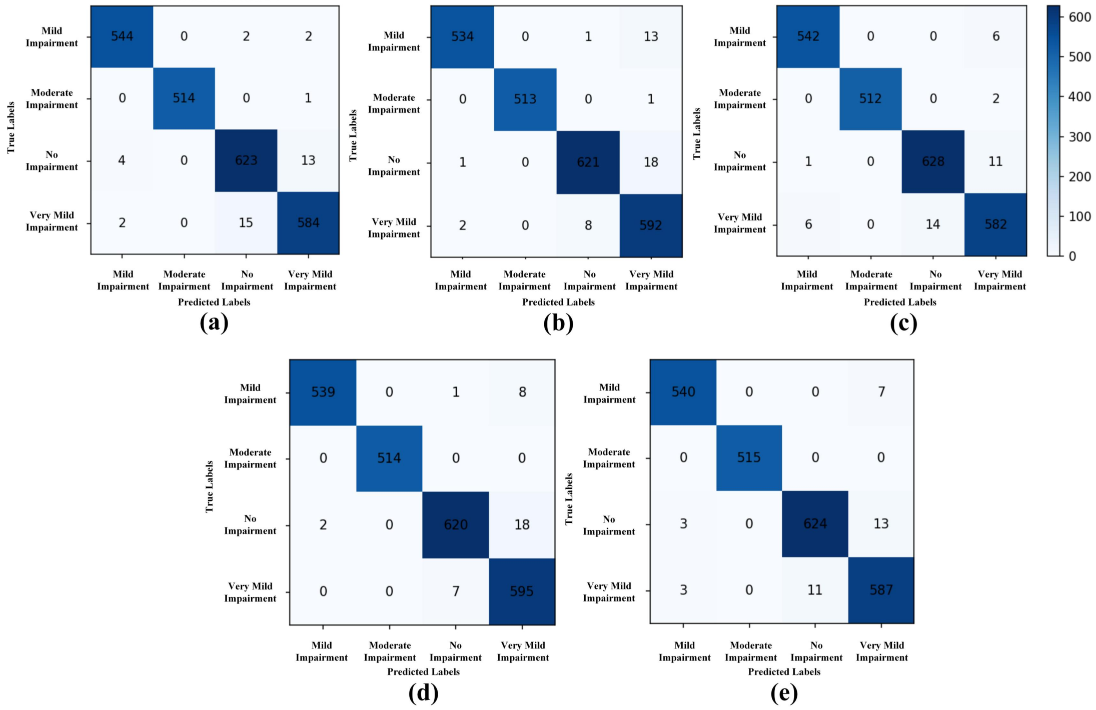

Figure 5: Confusion Matrix Diagram

To assess how well the proposed method was able to classify the various stages of Alzheimer's methodology in greater detail, we pulled several different metrics gathered from five-fold cross-validation analysis to provide detailed visuals (Confusion Matrices) on how well our model performed; See Sample Figure 5 for examples of confusion matrices representing each of the five folds of validation.

From the results presented in the confusion matrices, we can see that for all classifications of No Impairment, Very Mild Impairment, Mild Impairment, and Moderate Impairment, there is an overall strong diagonal dominance where the majority (91%) of the representatives of each classification accurately represent their respective categories; thus supporting that our model offers an accurate representation of how well it performed based on this trial group. A more detailed analysis of the confusion matrices shows that although the misclassifications of the subjects in one of the four classifications were not random in distribution, there was a predominance of such misclassifications among adjacent levels of impairment along the continuum of Alzheimer disease progression; as such, Very Mild Impairment and Mild Impairment as well as Mild Impairment and moderate Impairment have considerable degrees of diagnostic ambiguity due to clinical similarity between these two conditions. Thus, this diagnostic pattern serves to confirm the interdisciplinary nature of the clinical diagnostic process in the identification of both Very Mild and Mild Impairment and demonstrates that although there is some degree of differentiation between these two classifications of impairment, the absence of other co-morbidities greatly reduces one's diagnostic ability. In summary, the analysis of the confusion matrix demonstrates that not only does the model demonstrate overall strong performance across a wide variety of imaging modalities; but also provides insight into the confusion regions of our model's decision boundary when trying to accurately classify the various stages of Alzheimer's Disease using medical imaging technologies at the initial stages of Alzheimer's Disease.

3.5 Comparison with Different Models

The lasting benefit of EDA-Net and the implications of its design were determined by testing EDA-Net against several representative state-of-the-art models including ResNeXt50 [20], MobileNetV2 [24], DenseNet121 [25], and Vision Transformer (ViT) [26] on the same datasets and experimental setup. Table 2 shows that EDA-Net continuously surpassed all comparative methods using the same evaluation metrics, confirming that it produces superior overall performance and stability when compared to ResNeXt50 as the strongest performing model within the tested models with a difference of approximately 2.6 percentage points more accuracy than ResNeXt50 on the testing data. This increase in accuracy is due to implementing the proposed Efficient Channel–Spatial Attention Module (ECSAM), which allows for dynamic calibration of both inter-channel correlation as well as attention on key spatial locations, thereby enabling the ability to capture more accurately fine-grained local detail, as well as long-range spatial relationships due to the typical changing patterns seen in the early stages of developing Alzheimer's Disease (AD). MobileNetV2 and DenseNet121 also have comparable performances to EDA-Net; however, neither uses a mechanism designed to capture the local detail and long-range dependencies that characterize medical images, thus reducing overall performance when compared to EDA-Net. In comparison, ViT performed poorly and was extremely unstable throughout the test for EDA-Net. This is mainly because pure Transformer architectures rely heavily on large-scale pretraining to compensate for the absence of inherent image inductive bias, making them less effective when trained on moderately sized medical imaging datasets. Overall, through task-oriented architectural design, EDA-Net achieves state-of-the-art performance on the challenging task of multi-stage AD classification.

Table 2: Performance Comparison of Different Models on the AD Multi-Stage Classification Task

Models	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)	Kappa (%)
ResNeXt50[20]	95.72 ± 0.39	95.88 ± 0.38	93.49 ± 0.41	93.42 ± 0.41	94.28 ± 0.52
MobileNetV2[24]	93.25 ± 0.37	93.45 ± 0.35	93.60 ± 0.38	93.51 ± 0.37	90.97 ± 0.49
ResNet50[27]	93.14 ± 0.45	93.38 ± 0.40	93.49 ± 0.41	93.42 ± 0.41	90.83 ± 0.60
DenseNet121[25]	94.80 ± 0.33	95.00 ± 0.27	95.06 ± 0.26	95.02 ± 0.27	93.05 ± 0.45
EfficientNet[28]	94.42 ± 0.24	94.62 ± 0.21	94.69 ± 0.25	94.64 ± 0.22	92.54 ± 0.32
GhostNetV3[29]	89.77 ± 1.08	90.16 ± 1.00	90.23 ± 0.92	90.18 ± 0.96	86.31 ± 1.45
ViT[26]	47.32 ± 6.76	47.69 ± 7.10	47.32 ± 6.76	46.15 ± 7.64	29.90 ± 9.10
EDA-Net	98.30 ± 0.12	98.37 ± 0.13	98.39 ± 0.11	98.38 ± 0.12	97.73 ± 0.16

4. Conclusion

In this study, we propose a novel deep learning framework, termed EDA-Net, for accurate multi-stage classification of Alzheimer's disease (AD) based on brain magnetic resonance imaging (MRI). The core contribution of this framework lies in the design and implementation of a Deep Attention Block (DA Block) that deeply integrates dual attention mechanisms, as well as the construction of an end-to-end feature extraction network built upon this block. In order to meet the current demand for improved methods of detecting early AD using imaging techniques, we created an Efficient Channel-Spatial Attention Module (ECSAM). It includes two components: Efficient Channel Attention (ECA) and Coordinate Attention (CA). By sequentially combining these two techniques, we are able to increase the ability of our model to recalibrate how information flows through channels as well as provide more accurate targeting of spatial regions within the images we process. We utilised Group Normalization (GN) and SiLU activation in our DA Block architecture to improve the training stability and nonlinear representation capacity of our model in a small-batch context when working with medical imaging data.

Comprehensive cross-validations across five folds conducted on a publicly available, large-scale AD MRI dataset confirm the efficacy of EDA-Net from multiple angles. The findings of our ablation studies also support that the combination of both channel and coordinate based attention is warranted and leads to significant improvements in performance when combined together versus alone. The confusion matrix analysis indicates that the model makes errors that correspond to how AD progresses over time, which serves as additional validation for the representations we trained the model to learn. We performed side-by-side comparisons with many popular deep learning architectures, listed in Table 2, in order to demonstrate that EDA-Net consistently outperforms its peers across key measures (accuracy, recall, etc.). Specifically, we found that EDA-Net easily outperformed ResNeXt50 and DenseNet121 when distinguishing Very Mild impairment, which is an extremely challenging subset to identify in a clinical setting. This demonstrates the potential value of EDA-Nets architectural approach.

In summary, EDA-Net provides a high-accuracy and robust automated solution for early screening and fine-grained staging of Alzheimer's disease. Beyond advancing the application of deep learning in neuroimaging analysis, this method offers a powerful technical tool to support clinical decision-making and facilitate early intervention and disease management. Future work will focus on validating the generalization ability of the model on more diverse, multi-center clinical datasets and exploring multimodal information fusion to accelerate translation into real-world clinical practice.

References

[1] P. Scheltens *et al.*, "Alzheimer's disease," *The Lancet,* vol. 397, no. 10284, pp. 1577-1590, 2021.

[2] G. B. Frisoni *et al.*, "Alzheimer's disease outlook: controversies and future directions," *The Lancet,* vol. 406, no. 10510, pp. 1424-1442, 2025.

[3] C. R. Jack Jr *et al.*, "Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup," *Alzheimer's & Dementia,* vol. 20, no. 8, pp. 5143-5169, 2024.

[4] M. A. Better, "Alzheimer's disease facts and figures," *Alzheimers Dement,* vol. 19, no. 4, pp. 1598-1695, 2023.

[5] B. Dubois *et al.*, "Alzheimer disease as a clinical-biological construct—an International Working Group recommendation," *JAMA neurology,* vol. 81, no. 12, pp. 1304-1311, 2024.

[6] S. Aramadaka *et al.*, "Neuroimaging in Alzheimer's disease for early diagnosis: a comprehensive review," *Cureus,* vol. 15, no. 5, 2023.

[7] A. El-Assy, H. M. Amer, H. Ibrahim, and M. Mohamed, "A novel CNN architecture for accurate early detection and classification of Alzheimer's disease using MRI data," *Scientific Reports,* vol. 14, no. 1, p. 3463, 2024.

[8] M. Charquero-Ballester *et al.*, "Effective psychological therapy for PTSD changes the dynamics of specific large-scale brain networks," *Human brain mapping,* vol. 43, no. 10, pp. 3207-3220, 2022.

[9] A. W. Salehi *et al.*, "A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope," *Sustainability,* vol. 15, no. 7, p. 5930, 2023.

[10] N. Ajlouni, A. Özyavaş, M. Takaoğlu, F. Takaoğlu, and F. Ajlouni, "Medical image diagnosis based on adaptive Hybrid Quantum CNN," *BMC Medical Imaging,* vol. 23, no. 1, p. 126, 2023.

[11] M. M. Musthafa, M. TR, V. K. V, and S. Guluwadi, "Enhanced skin cancer diagnosis using optimized CNN architecture and checkpoints for automated dermatological lesion classification," *BMC Medical Imaging,* vol. 24, no. 1, p. 201, 2024.

[12] B. Khagi and G.-R. Kwon, "3D CNN design for the classification of Alzheimer's disease using brain MRI and PET," *IEEE Access,* vol. 8, pp. 217830-217847, 2020.

[13] G. Mohi ud din dar *et al.*, "A novel framework for classification of different Alzheimer's disease stages using CNN model," *Electronics,* vol. 12, no. 2, p. 469, 2023.

[14] W. Salehi *et al.*, "An approach to binary classification of Alzheimer's disease using LSTM," *Bioengineering,* vol. 10, no. 8, p. 950, 2023.

[15] A. A. A. El-Latif, S. A. Chelloug, M. Alabdulhafith, and M. Hammad, "Accurate detection of Alzheimer's disease using lightweight deep learning model on MRI data," *Diagnostics,* vol. 13, no. 7, p. 1216, 2023.

[16] X. Zhang, L. Gao, Z. Wang, Y. Yu, Y. Zhang, and J. Hong, "Improved neural network with multi-task learning for Alzheimer's disease classification," *Heliyon,* vol. 10, no. 4, 2024.

[17] A. S. Alatrany, W. Khan, A. Hussain, H. Kolivand, and D. Al-Jumeily, "An explainable machine learning approach for Alzheimer's disease classification," *Scientific reports,* vol. 14, no. 1, p. 2637, 2024.

[18] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11534-11542.

[19] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 13713-13722.

[20] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1492-1500.

[21] Y. Wu and K. He, "Group normalization," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3-19.

[22] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," *arXiv preprint arXiv:1710.05941,* 2017.

[23] L. Chugh, O. Alshabrawy, and Y. Shang, "Addressing Data Scarcity and Class Imbalance in Alzheimer's," *Nternational Journal of Engineering Research & Technology (IJERT),* pp. I-40, 2023.

[24] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510-4520.

[25] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700-4708.

[26] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929,* 2020.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[28] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, 2019: PMLR, pp. 6105-6114.

[29] Z. Liu, Z. Hao, K. Han, Y. Tang, and Y. Wang, "Ghostnetv3: Exploring the training strategies for compact models," *arXiv preprint arXiv:2404.11202,* 2024.