INTERNATIONAL JOURNAL OF
Telecommunications
AIR DEFENSE COLLEGE

Elevating Alzheimer's Diagnosis based on Attention-Guided MRI Feature Fusion

Elassy, A. M.[1] *, Amer, Hanan. M.[2], Ibrahim, H.M.[3], and Mohamed, M.A.[4]

[1] Electronics and Communications Engineering Department - Faculty of Engineering – Mansoura University – Mansoura – Egypt; ayah_elassy@mans.edu.eg.
[2] Electronics and Communications Engineering Department - Faculty of Engineering – Mansoura University – Mansoura – Egypt; hanan.amer@yahoo.com.
[3] Assistant Professor- Faculty of Computers and Information Systems -Egyptian Chinese University-Cairo - Egypt; hegazibrahim@gmail.com.
[4] Electronics and Communications Engineering Department - Faculty of Engineering – Mansoura University – Mansoura – Egypt; mazim12@mans.edu.eg.

Citation: Elassy, A.M.; Amer, Hanan. M., Ibrahim, H.M., Mohamed, M.A.
*Inter. Jour. of Telecommunications, **IJT'2025**, Vol. 05, Issue 01, pp. 01-20, 2025.*

Editor-in-Chief: Youssef Fayed.

Received: 05/01/2025.

Accepted: 19/02/2025.
Published: 19/02/2025.

Abstract: The complicated pathophysiology of Alzheimer's disease (AD), which can occasionally be inherited, is typified by the loss of synapses and neurons as well as the appearance of neurofibrillary tangles and senile plaques. For treatment or prevention to be effective, early detection is essential, especially in high-risk patients. This work offers a multi-model feature fusion method based on the attention mechanism as a novel way to classify Alzheimer's disease. The ADNI dataset was first used to test many pre-trained models, and the top three performances were chosen for additional testing. We created an attention-based feature fusion module to efficiently combine features from three different modalities. Our tests showed that merging features without the attention mechanism results in a significant decline in performance (accuracy=82%). However, implementing the attention mechanism before the fusion process significantly enhanced performance, with 99.31% accuracy in classifying Alzheimer's disease into five stages. Motivated by these outcomes, we expanded our approach to classify the disease into four and three stages, with 98.29% and 99.43% accuracies, respectively. Our results demonstrate how well the multi-model features with the attention mechanism work to improve Alzheimer's disease classification.

Keywords: Alzheimer's disease, deep learning, attention mechanisms, MRI.

1. Introduction

AD is a neurological illness that worsens over time[1] . It is thought that abnormal protein buildup within and around brain cells causes AD. When clinical signs are paired with histological evidence of amyloid plaques and the kind of tau tangles in post-mortem brain tissue, AD cannot be conclusively identified until death [2]. Classifying AD has primarily focused on neuroimaging in recent years, with magnetic resonance imaging (MRI) being a particularly popular method [3]. There are five stages of Alzheimer's disease: (AD), Cognitively Normal (CN), Early Mild Cognitive Impairment (EMCI), Mild Cognitive Impairment (MCI), and Late Mild Cognitive Impairment (LMCI). Alzheimer's symptoms in their early stages include forgetting recent conversations or events, losing things, forgetting locations and objects, struggling to think of the proper phrases, asking the same questions over and over, and showing poor self-control. In AD, brain scans identify alterations brought on by brain cell death [4]. To make training the CNN architecture's deep learning technique easier, the image is preprocessed. Convolutional and pooling layers are used by convolutional neural networks (CNNs), which are specialized deep learning models for image processing, to automatically learn hierarchical features.

CNN offers greater accuracy [5, 6]. Deep learning methods offer a lot of promise for disease classification[7] . The extraction of both simple and complicated features is made easier by deep learning architectures, which are made up of several layers of neurons stacked hierarchically. Higher layers are skilled at identifying more complex patterns, whereas the lowest layers are focused on capturing basic traits [8]. There are several parameters in each layer that need to be adjusted during the training process. In order to minimize the selected loss function—such as l2 loss, cross-entropy loss, or other variations—back-propagation is typically used during training. Long training timeframes result from deep learning networks' frequent inability to converge to the global optimum, even with sophisticated optimization strategies.

This study investigates how deep learning techniques might enhance AD classification with just MRI data. To automatically extract complex information from MRI scans and categorize people into various phases of AD, we provide a unique approach that makes use of a deep learning architecture. The suggested technique seeks to identify complex patterns and relationships in the MRI data to get over the drawbacks of innovative machine learning techniques. This study's primary contributions are as follows:

- Creative Method: The study presents a brand-new feature fusion module that greatly improves Alzheimer's disease classification performance. It is based on the attention mechanism.
- Extensive Analysis: The study offers a comprehensive comparison of different fusion techniques and pre-trained models, emphasizing the role that the attention mechanism plays in increasing accuracy.
- Enhanced Diagnostics: The suggested method provides important insights and possible uses in early diagnosis and therapy planning by attaining high accuracy in classifying Alzheimer's disease into different phases.
- Evaluating the model rigorously: The study uses a robust evaluation methodology that includes ROC curves, confusion matrices, accuracy, recall, and specificity.
- Comparison with state-of-the-art techniques: Using solely MRI data, the suggested method's superiority in classification accuracy is demonstrated by comparison with current techniques.

The article is further structured as follows: The first section presents the relevant research on the early identification of AD. The study's data collection and preparation are covered in the second section. The creation of the suggested methodology makes up the third section. In the fourth section, the overall experiments and results are presented. Its debates are listed in the fifth section. The sixth section provides a summary of this study.

2. Literature Review

Significant progress has been made in the field of medical imaging in recent years, particularly in the diagnosis and detection of brain disorders like AD [9]. as well as MCI. Techniques for Deep Learning (DL) and Machine Learning (ML) have been essential in improving the precision and effectiveness of diagnostic systems for various illnesses [10]. This review of the literature examines the different approaches and strategies used in current research on the identification and diagnosis of AD and MCI, with a focus on the application of both conventional machine-learning techniques and deep learning approaches. A refined CNN architecture was created by Shamrat et al. [11] to forecast Alzheimer's disease in five phases. Depending on the layers and hyperparameters, they alter the model. Then, using 60,000 MRI scans from the ADNI database, they achieved the greatest accuracy of 96.31%. Tanveer et al. [12] introduced a computationally effective neural network ensemble that was trained via transfer learning. The classification accuracy for AD and MCI was 83.11% and 98.71%, respectively, on two separate datasets that were divided by AD and NC. A Siamese Convolutional Neural Network (SCNN) architecture was created by Hajamohideen et al. [13] that embeds input images as k-dimensional embeddings with a triplet loss function. Using both pre-trained and untrained CNNs, Alzheimer's illness was classified using this embedding space. Using the ADNI and OASIS datasets, the model's efficacy was assessed;

the accuracy rates were 91.83% and 93.85%, respectively. Using feature selection methods and supervised learning algorithms, Divya et al. [14] classified AD using MRI features from the ADNI dataset. A support vector machine (SVM) using a radial basis function kernel produced the greatest results, with 96.82%, 89.39%, and 90.40% accuracy for the binary classifications of NC/AD, NC/MCI, and MCI/AD, respectively. To maintain low-intensity pixels, this study [15] modified the LeNet model by concatenating Min-Pooling layers with Max-Pooling layers. When compared to 20 other DNN models, this new model outperformed the others, with an average classification accuracy of 96.64% for Alzheimer's disease vs 80% for the original LeNet.

A feature mining technique based on deep learning has been proposed by Sharma et al.[16] Following feature extraction, the AD was classified using a "Fuzzy Hyperplane Least Square Twin Support Vector Machine (FLS-TWSVM)." They have used the sagittal plane slices of the 3D MRI image. This investigation used the ADNI dataset. A hyperplane was created using the triangular fuzzy function to carry out classification. It was shown that the proposed model's classification accuracy for AD was higher than the other models already in use. To effectively classify AD, Kumar et al. [17] used an AlexNet model to identify the most noticeable features from the input MRI scans. At the MCI level, this model was able to diagnose AD. The created framework was tested using a vast amount of MRI images from the "Open Access Series of Imaging Studies (OASIS) Brain" dataset. Comparing the proposed model to its predecessors, the accuracy was higher. Using the Adam optimizer algorithm, Borkar et al. [18] proposed combining CNN and LSTM to identify AD in its early phases. The MRI scans were used to extract the brain's characteristics. With 99.7% improved detection accuracy, this approach was economical. Muhammad et al. [19] classified AD from MRI scans using a deep CNN model. This model produced a thermal map of the brain. This model accurately distinguished between the early phases of AD for a small sample. This model offered a low-cost classification model and was constructed with fewer parameters. An oversampling technique was used to address the dataset's imbalance problems. When it came to categorizing the early phases of AD, our suggested model performed better than other methods now in use. Sharma et al. [20] have created a method to identify AD in its early stages. A neural network was used to construct the deep learning model. The VGG-16 model was utilized to extract the characteristics. The created neural network was used to detect AD using two different MRI datasets with many sample images. The results of the simulation demonstrated that the proposed neural network was more effective at identifying AD in its early stages.

Numerous researchers have put forth novel strategies for the early detection of AD by applying deep learning methods to medical imaging data, including MRI scans. A CNN-SVM model, which combines the feature extraction skills of CNN with the classification capabilities of SVM, was presented by Sethi et al. [21]. With an astounding accuracy of 86.2% on the OASIS dataset, this model demonstrated relative accuracy gains ranging from 0.85 to 3.4% on various datasets. The model has demonstrated the ability to diagnose this illness with high accuracy and performs exceptionally well with complex datasets. These benefits may be essential for early AD diagnosis and for taking into account additional research in the specific field. [22]used multimodal neuroimaging data to develop a diagnostic method for AD. This approach uses a cutting-edge zero-masking technique that maintains all the data and the information it contains. SAE is used to extract high-level features, which are then fed into SVM for multi-modal and multi-class classification of MR/PET data. The study found that the model performed with an accuracy of 86.86%, suggesting that the used strategy may be useful for AD early identification. For learning nonlinear features, the unsupervised autoencoder technique is helpful, but it has drawbacks, such as model overfitting and problems with interpretability and generation. The SAEs may be difficult to interpret due to their intricacy, which could prevent clinics from using them where interpretability is crucial. Using a 3D-UNET architecture, Ruchika et al. created a robust classification system for AD [23] stages through volumetric MRI data analysis, achieving impressive results with a segmentation train accuracy of 93% and a test accuracy of 90%. They also used a novel volumetric analysis approach to classify the four phases of AD based on hippocampus volume, achieving an accuracy of 91% person-wise and 88% hemisphere-wise after modifying the threshold using the root mean square error (RMSE). To categorize Alzheimer's disease, Manop et al. employed Deep transfer learning models with oversampling [24]. The researchers classified the stages of Alzheimer's disease progression using deep transfer learning models and oversampling techniques. The transfer learning models employed in the study were MobileNetV2, ResNet50, Xception, and VGG19.

With Xception topping the pack at 82.46%, MobileNetV2 coming in second at 79.29%, VGG19 at 77.73%, and ResNet50 at 76.28%, the models' accuracy varied. A dataset of 6,327 occurrences divided into four groups was used in the study. The results of this study provide insight into how well deep transfer learning models classify the course of Alzheimer's disease, with Xception emerging as the most accurate model.

The application of Vision Transformer (ViT) to MRI processing for Alzheimer's disease diagnosis was examined by Alp et al. [25]. After the MRI features were retrieved and modeled, ViT was utilized as a time series transformer to classify them. The model was tested for binary and multiclass classification on ADNI T1-weighted MRIs. With scores exceeding 95% for binary and 96% for multiclass classification, the model showed a high degree of accuracy when compared to deep learning models like CNN with BiL-STM and ViT with Bi-LSTM. Shaffi et al. [26] provided a machine learning-based ensemble classifier for MRI scan-based AD prediction. With a remarkable 96.52% accuracy rate, it outperformed the top individual classifier by 3–5%. Using the Open Access Series of Imaging Studies and Alzheimer's Disease Neuroimaging Initiative datasets, they evaluated popular machine learning classifiers and achieved an enhanced accuracy of over 94%. This paper [27] proposes a Bi-Vision Transformer (BiViT) architecture for categorizing various cognitive diseases and AD phases from 2-dimensional MRI imaging data. To be more precise, the transformer is made up of two innovative modules for efficient feature learning: Mutual Latent Fusion (MLF) and Parallel Coupled Encoding Strategy (PCES). The performance of the suggested BiViT-based architecture has been assessed using two distinct datasets. The findings demonstrate that, when applied to the AD dataset, the suggested BiViT-based model attains an accuracy of 96.38%. Nevertheless, the accuracy drops slightly below 96% when applied to data on cognitive diseases, which may be the consequence of an unbalanced distribution of data and a lesser amount of data. A bilateral filtering and histogram equalization picture enhancement approach was proposed in the study [28] to improve the dataset's quality. After that, a customized CNN architecture was created in order to categorize dementia into three categories. With the help of the specially created architecture, the correctness of the provided architecture was 93.45% for the multiclass and 95.62% for the binary class.

To predict the development of Alzheimer's disease, the suggested method [29] integrates single-nucleotide polymorphisms (SNPs), ratio of grey matter volume (RGV), and structural MRI (sMRI) images using a trimodal data fusion approach. Among the study's significant findings is the 94.37% classification accuracy attained in differentiating between progressive mild cognitive impairment (pMCI) and stable mild cognitive impairment (sMCI), which surpasses current state-of-the-art techniques. The method's superior predictive performance and ability to use multiple modalities for increased accuracy are its advantages; however, the participants' educational backgrounds and the requirement for additional validation across a variety of datasets could potentially introduce biases. Using structural MRI and fMRI data from the ADNI dataset, the study [30] uses a hybrid CNN-RNN model, emphasizing strict data preprocessing and quality control to improve model performance. The model's outstanding performance in multiclass categorization of AD and MCI is demonstrated by key results that show strong classification accuracies of 99.5% in testing, 97.2% in validation, and 94.0% in real-world settings. This method's benefits include increased diagnostic precision, the possibility of early diagnosis, and customized treatment plans; its drawbacks include the need for high-quality data.

3. Materials

3.1. Description of the AD Dataset

This study makes use of the Alzheimer's Disease Neuroimaging Initiative's MRI ADNI dataset [31]. The ADNI complied with the Health Insurance Portability and Accountability Act and was authorized by the Institutional Review Board at each of the ADNI Clinical Trial Centers. Before beginning the trial, each patient gave their informed consent.

As shown in Fig. 1, this dataset consists of 1,296 T1-weighted MRI scans with 1.5 mm isotropic voxel resolution that are divided into five classes: AD, MCI, EMCI, LMCI, and CN. The dataset has significantly more CN than the other classifications, indicating a statistical imbalance. Participants' average ages range from 69.3 years for EMCI to 76.2 years for AD, and their average educational attainment is roughly 16 years for all classes. There are notable variations between the classes in important neuroimaging biomarkers, including mean diffusivity (MD) of white matter, fractional anisotropy (FA) of the corpus callosum, and hippocampus volume. Neurodegenerative alterations are evident in AD patients, who have the highest average MD (0.85) and the lowest average hippocampus volume (4.5 cubic centimeters). Conversely, CN has the lowest average MD (0.75) and the largest average hippocampus volume (6.0 cubic centimeters).

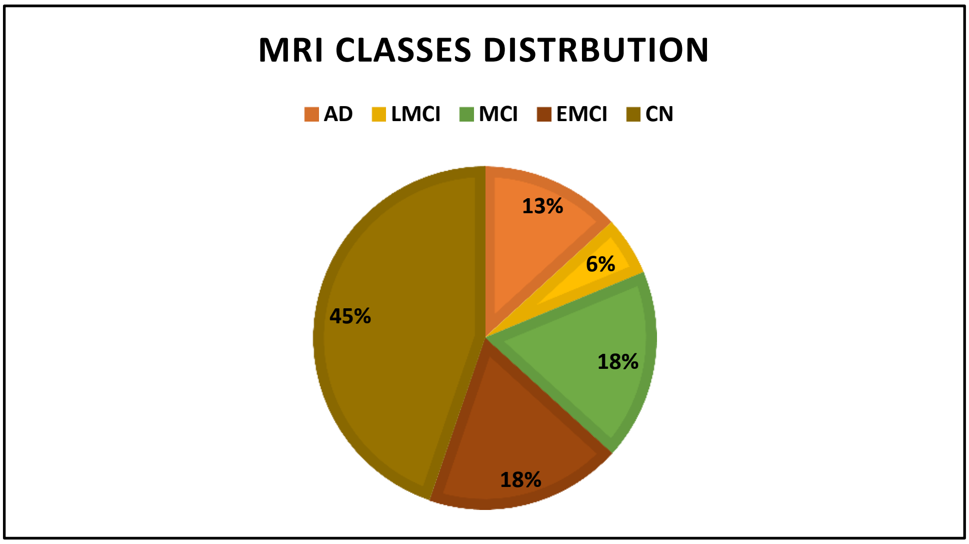


Figure 1 class distribution **of** the MRI dataset

3.2 Data Preprocessing

ADNI-funded MRI Analysis Laboratories processed and prepared the data. The scenes in the dataset were subjected to necessary pre-processing procedures to maximize their quality and consistency for analysis. Among these actions were:

- **Scaling**: All scenes are uniformly resized to 224 pixels in height and width.
- **Augmentation**: Using data augmentation approaches, as mentioned in [32-34] to improve the dataset's variety and reduce overfitting.
- **Balancing the AD Dataset**: We created artificial data for under-represented classes using the ADASYN technique to address the class imbalance problem shown in Figure 1. By adaptively creating synthetic samples according to the difficulty of learning each minority class instance, ADASYN, an advanced oversampling technique, gets beyond the drawbacks of SMOTE. ADASYN reduces bias and enhances model performance by deliberately concentrating on difficult samples. As illustrated in Figure 2, ADASYN lowers the danger of overfitting and improves the overall representation of minority classes by introducing new data points in areas where they are most needed, in contrast to SMOTE, which repeats existing minority class points [35].
- **Data Splitting**: Three subsets of the dataset were created: testing, validation, and training. 10% of the data was allocated to the test set, while 90% of the data was randomly assigned to the training set. For the training set, cross-validation was used to reduce the possibility of dataset bias. With this method, the training data is divided into several subsets, and the performance of the model is evaluated using each subset as a validation set. For hyperparameter tweaking, including regularization and learning rate, the validation set is essential to avoiding overfitting and improving accuracy. Training is stopped once the model converges on the validation set in order to prevent repeating experiments. Accuracy, precision, recall, and other measures are used to evaluate the final model's generalization capacity on the independent test set.

4. Methodes

4.1 Deep Learning Architecture

To create a hybrid deep learning architecture, a performance comparison was conducted among different types of pre-trained models (CNN, DenseNet121, ResNet50v2, MobileNet, NASNetMobile, AlexNet) after applying them to a dataset as shown in Table 1, and selecting the modules with the best performance, which are DenseNet-121, ResNet-50v2, and Custom CNN[5]

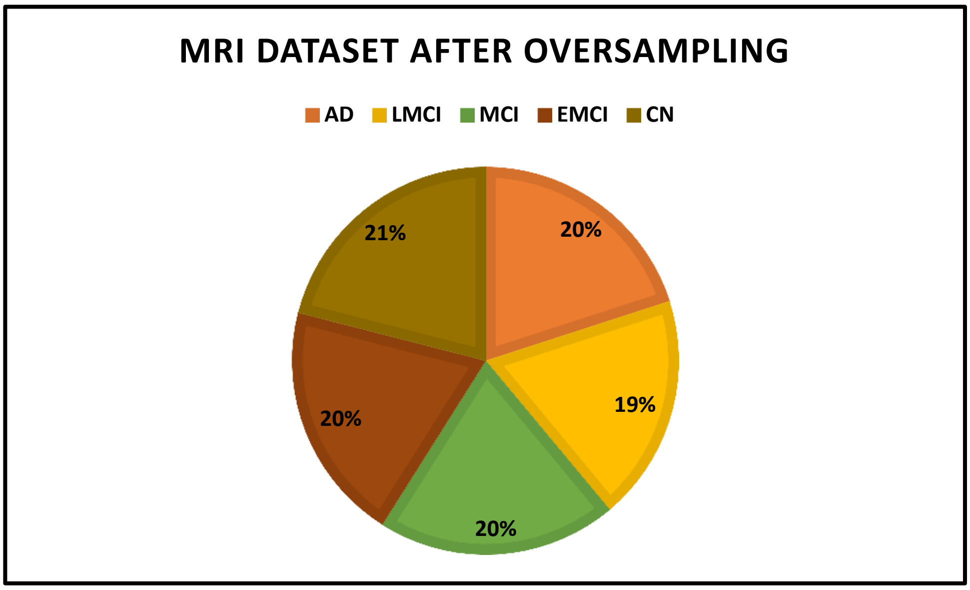


Figure 2 class distribution of the MRI dataset after oversampling.

Table 1 performance comparison between deep learning architectures

Models	Loss	Accuracy	Precision	Recall
DenseNet121	0.1596	95.54%	95.54%	95.54%
Custom CNN	0.3286	95.42%	95.74%	95.07%
ResNet50v2	0.1898	94.39%	94.72%	94.39%
MobileNet	0.1533	93.54%	95.19%	92.62%
NASNetMobile	0.2013	93.35%	94.23%	92.35%
AlexNet	0.3095	90.10%	91.78%	88.16%

4.2 The Proposed Model Description

The proposed method involves combining the results of three different models: Resnet50 v2, Densenet121, and a custom CNN model that is characterized by its simplicity and potential for heightened performance. As presented in Table 1. These choices were taken after a careful analysis in which the models demonstrated their ability to handle complex tasks and outperform their rivals. The custom CNN architecture includes a sequence of convolutional and pooling layers that make up the bespoke CNN model that is suggested in this work. Low-level features are captured by the first 16-filter convolutional layers, whereas more intricate patterns are extracted by the next 64- and 256-filter layers. The feature maps are down-sampled by the max-pooling layers, which lowers computing costs and boosts resilience to slight input fluctuations. To enhance generalization and stabilize the training process, batch normalization is used to create a precisely customized feature extraction pipeline.

Before the merging of these models, each model's output is subjected to a custom soft attention process that sharpens the focus on important details and is skilled at focusing attention on input elements, increasing the model's ability to efficiently extract important information. This is followed by global average pooling to condense spatial information, culminating in a consolidated feature representation.

The model's capacity to generalize is strengthened, and overfitting is avoided by incorporating a fully connected layer with 1024 neurons and 50% dropout for regularization. Softmax activation in the classification layer provides the crucial output for the multiclass classification job. To fully realize the potential of this complex

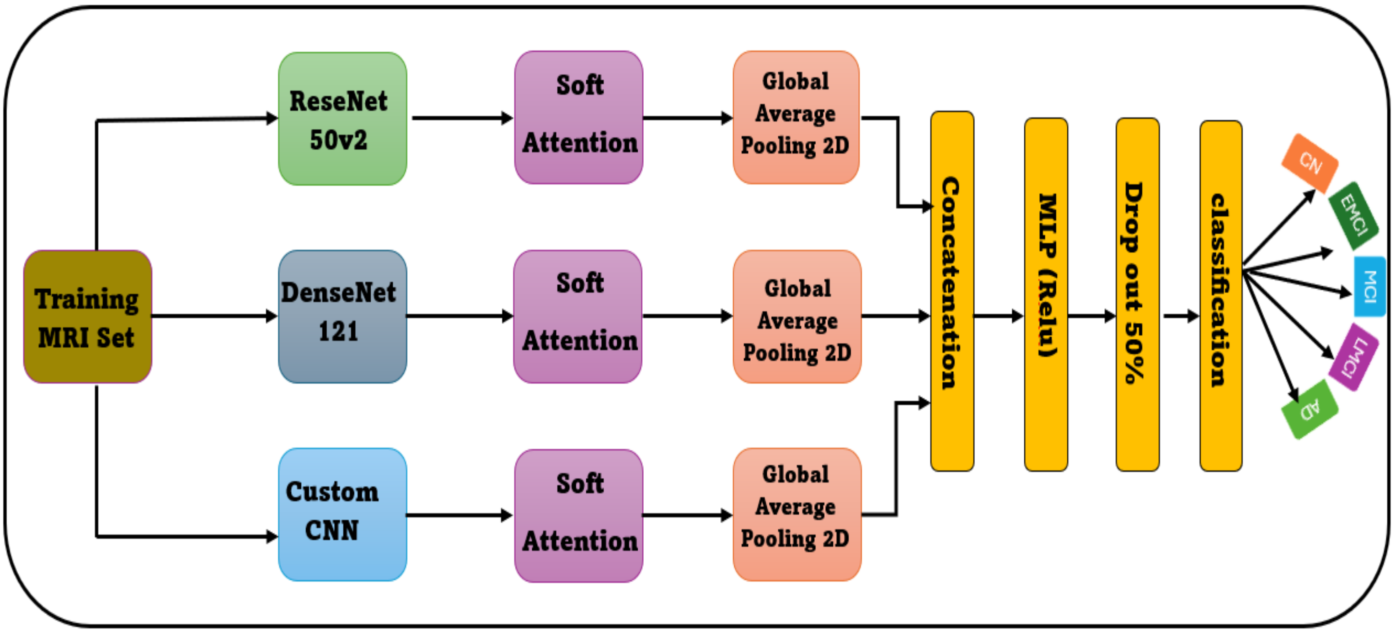


Figure 3 block diagram of proposed Model

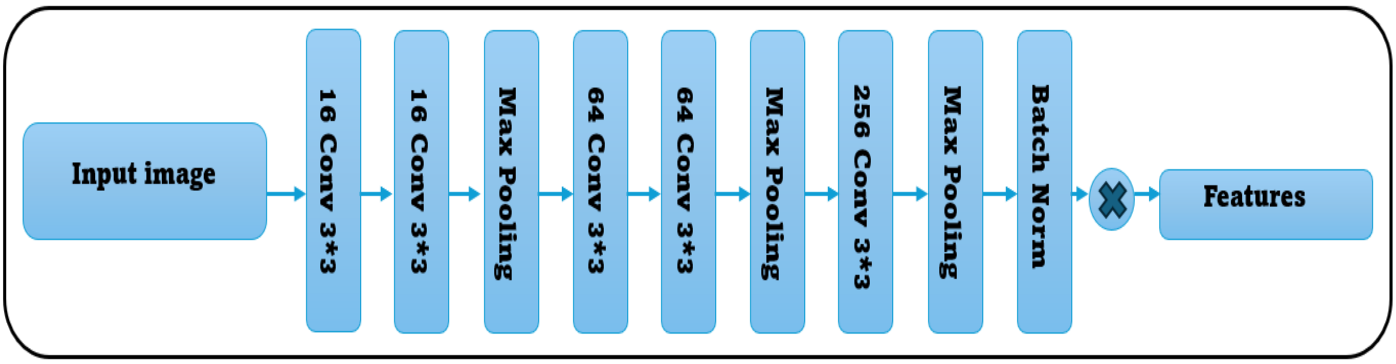


Figure 4 block diagram for custom CNN (Cited by [5])

network, careful optimization and fine-tuning are necessary, even though the suggested design unifies a wide range of models and methodologies, utilizing their advantages to create a more robust and accurate categorization system. Figure 3 provides a comprehensive visual representation of the information discussed so far. In the following sections, we will examine each part of this figure in greater detail.

4.2.1 The Residual Network ResNet50V2

Microsoft Research unveiled ResNet50, a 50-layer deep convolutional neural network, in 2015. It effectively trains very deep networks by addressing the vanishing gradient issue with residual learning and skipping connections. Compared to shallower models like VGG, the architecture's bottleneck blocks improve accuracy and processing efficiency.

The initial ResNet50 model served as the basis for the more advanced ResNet50v2 convolutional neural network design, which addressed the vanishing gradient problem in deep networks in a novel way by using residual connections. ResNet50v2, created by Kaiming He and colleagues, has several enhancements designed to increase training effectiveness and deep learning model performance. designed especially to allow for deeper networks without sacrificing performance[36-38].

4.2.2 Densely Connected Convolutional Network DenseNet 121

The convolutional neural network (CNN) architecture DenseNet-121 is renowned for its dense interlayer connections. Due to the efficient feature propagation and reuse made possible by this dense connectivity, performance is enhanced, and the number of parameters is decreased. In a dense block, every layer is intimately connected to every layer that came before it, enhancing feature propagation and allowing the network to extract more complex patterns from the data. Transition layers help manage model complexity by reducing the size of the feature map between dense blocks.

By lowering the quantity of input features at each layer, bottleneck layers increase computing efficiency. Each layer in a dense block adds a certain number of new feature maps, which is determined by the growth rate. A greater growth pace results in a more profound and intricate.[39]

4.2.3 Custom CNN Model

Convolutional Neural Networks work by gradually identifying characteristics in the input data. Simple elements such as corners and edges are recognized in the first layers. These characteristics are used to create increasingly intricate patterns as the network gets deeper, which eventually results in high-level representations[5].

CNNs are based on convolutional processes, which apply filters to the input data. These filters, also known as kernels, are parameters that can be trained to recognize patterns. A feature map, which shows whether the learned pattern was present in the input, is the result of a convolutional layer. Pooling layers are used to decrease the feature maps' spatial dimensions and add invariance to minor translations and distortions. Max pooling and average pooling are common pooling strategies. Another essential part of CNNs is batch normalization, which normalizes each layer's activations to stabilize training and speed up convergence. Lastly, the learned features are mapped to the output space, such as regression values or class probabilities, using fully connected layers.

As depicted in Figure 4. The custom CNN architecture includes 16 filters of size 3x3, subsequent max-pooling layers, followed by an additional two convolutional layers with 64 filters each, subsequent max-pooling layers, and culminating in a convolutional layer with 256 filters alongside batch normalization. The network can efficiently train discriminative representations from the input data thanks to the robust feature extraction foundation provided by this thoughtfully constructed architecture.

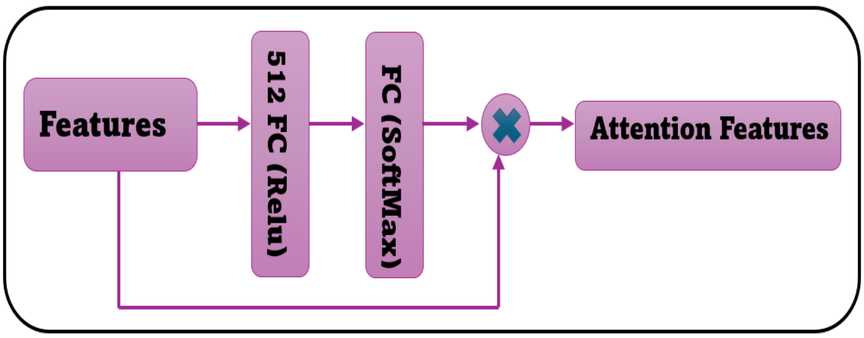


Figure 5 block diagram of Soft Attention

4.2.4 Soft Attention

A deep learning technique called "soft attention" enables neural networks to concentrate on areas of an input image while classifying it. It functions by assigning various areas of the image weights, with the most pertinent areas receiving the highest weights. With the aid of this weighted attention strategy, the network learns more discriminative features and increases classification accuracy.

According to figure 5 Soft attention is put into practice by assigning attention scores to each area of the image. Using these scores, a weighted total of the feature maps at each site is then produced. The network can then concentrate on the most instructive areas of the image by feeding the weighted feature maps into further layers.

In tasks like object detection and scene interpretation, researchers have seen impressive improvements by integrating soft attention into image classification algorithms. The way that various aspects of the image are ranked according to their significance to the job at hand is a method that replicates human visual attention. Thus, in addition to improving performance, soft attention offers insights into how neural networks may efficiently receive and interpret visual input, resulting in more reliable and understandable image classification systems[40].

4.3. Performance Evaluation

Each participant underwent stratified tenfold cross-validation as part of the model validation procedure. The batch size was 32, the learning rate was 0.001, and the model was trained for 50 epochs. To test the applicability of the top-performing model, we evaluated it on a cross-validation set. The model was assessed using performance metrics in both cross-validation studies, including accuracy, recall, specificity, Matthews Correlation Coefficient (MCC), and F1-score [41, 42]. Through the computation of the mean outcomes of ten-fold cross-validation, we assessed the model's overall effectiveness. The formulas below provide certain metrics' calculating methods:

$$\boldsymbol{accuracy = \frac{TP + TN}{TP + TN + FP + FN}} \quad (1)$$

$$\boldsymbol{Balanced\ ccuracy = \frac{Recall + Spesificity}{2}} \quad (2)$$

$$\boldsymbol{Recall = \frac{TP}{TP + FN}} \quad (3)$$

$$\boldsymbol{precision = \frac{TP}{TP + FP}} \quad (4)$$

$$\boldsymbol{MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}} \quad (5)$$

$$\boldsymbol{F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}} \quad (6)$$

These metrics primarily rely on components like True-Negative (TN), False-Negative (FN), True-Positive (TP), and False-Positive (FP). The number of samples that the network accurately identifies is shown by (TP) and (TN), which shows how well the model avoids false positives. However, samples that are wrongly identified are denoted by (FN) and (FP); a high rate indicates that the model may be excessively sensitive.

Accuracy is a measure of the classifier's capacity to distinguish across the whole sample. The classifier's recall indicates its capacity to distinguish between positive samples. The classifier's specificity indicates its capacity to distinguish between negative samples. The binary classification model's accuracy is evaluated using the F1 score, a statistical metric that takes into account both the model's recall and accuracy rates. The diagnostic accuracy of the classifier increases with the value. while The MCC is a more complex metric considering the dataset's imbalance between positive and negative examples.

Additionally, the Receiver Operating Characteristic (ROC) curve is plotted to evaluate the impact of various modules on the relationship between sensitivity and false positive rates. To give a thorough assessment of each module's contribution and the model's overall performance, the area under the ROC curve (AUC) is computed. The classifier's classification effect is intuitively assessed by the AUC value[43].

4.4. Model development and training

In our work, we used the Google Collaboratory Pro platform and Python 3.0 to train and evaluate the classifier [44]. A Tesla K80 GPU with 2496 CUDA cores, a computational capacity of 3.7, and 12 GB of GDDR5 VRAM was made available through this platform.

5. Experiments and Results

Several experiments were presented in this section. First, we used the ADNI dataset on various pre-trained model models. Three of the best-performing models were chosen based on Table 1. We created a multi-modal feature fusion module based on the attention mechanism in order to efficiently combine features from the three modalities. Next, as shown in Table 2, we experimented with contrasting the effects of feature fusion with and without the attention mechanism.

Table 2 performance enhancement of 3 bases model by combining features from three modalities and adding the attention mechanism

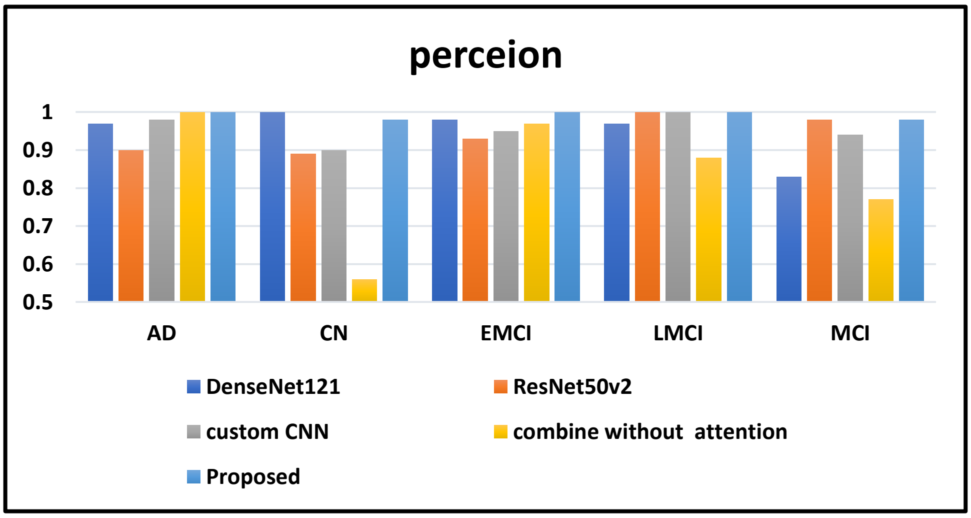
Metrics	DenseNet121	Resnet50v2	Custom CNN	Combining without attention	Combining with attention
Loss	0.1596	0.1898	0.3286	0.6344	**0.0219**
Accuracy	95.54%	94.39%	95.42 %	82.64%	**99.31%**
Precision	95.54%	94.72%	95.07 %	83.33%	**99.31%**
Recall	95.54%	94.39%	95.74 %	81.60%	**99.31%**
Balanced Accuracy	95.44%	94.72 %	95.45 %	82.56 %	**99.26 %**
MCC	95.12%	93.18 %	94.29 %	78.82 %	**99.13%**

Table 2 makes it clear that performance drastically declines when characteristics are combined without the attention mechanism. Where accuracy reaches 82.64%, while accuracy in each model reaches up to 95.54% in the case of DenseNet121, 97.39% in ResNet50v2, and 95.42% in custom CNN. Nevertheless, performance significantly improves when the attention mechanism is used before the fusion process, as seen in Figure 3. When Alzheimer's illness was divided into five stages, the classification accuracy hit 99.31%. This outstanding performance motivated us to use the suggested approach to categorize the disease into three and four stages, assessing accuracy in each instance. The accuracy was 98.29% for the four phases and 99.43% for the three stages, as shown in Table 3. These astounding outcomes demonstrate how well multi-modal features work with the attention mechanism to improve the classification of Alzheimer's disease.

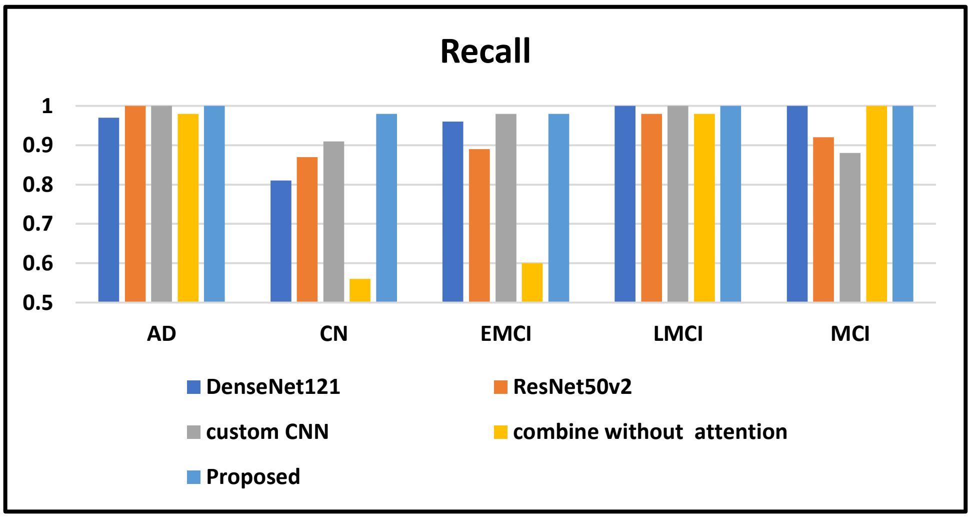
Table 3 performance of proposed network with 5-way multiclass; 4-way multiclass; and 3-way multiclass.

Metrics	5-Way Multiclass	4-Way Multiclass	3-Way Multiclass
Loss	0.0219	0.0693	**0.0137**
Accuracy	99.31%	98.29%	99.43%
Recall	99.31%	98.29%	99.26%
Precision	99.31%	98.29%	99.26%
Balanced Accuracy	99.26 %	98.46 %	99.32 %
MCC	99.13%	97.73 %	99.13 %

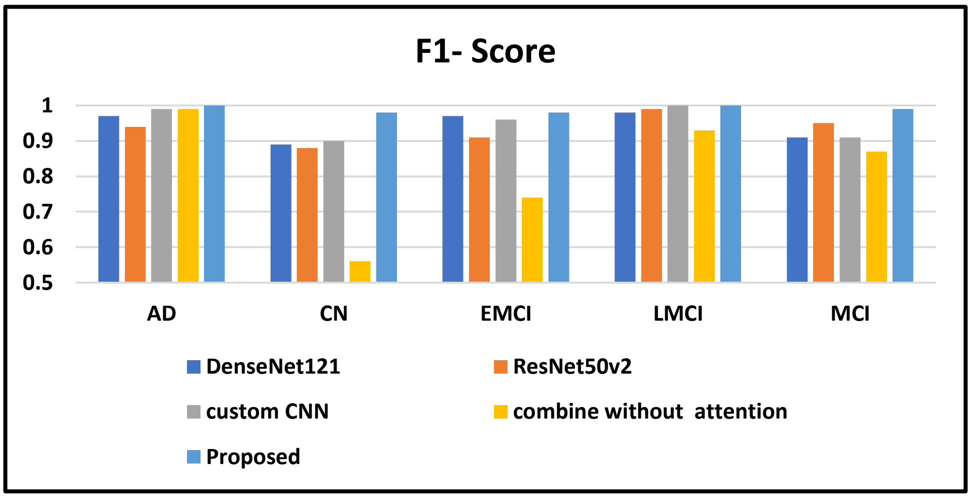
As seen in Figure 6, the F1-score, recall, and precision performance parameters for each class have improved because of the attention mechanism integration. The performance of each model separately and the suggested system after combining features from three modalities and adding the attention mechanism is contrasted in the figure. Figure 6's visualization makes it evident how the attention mechanism improves each class's performance on a variety of criteria. The significance of selective feature integration and attention mechanisms in boosting the accuracy and resilience of the classification system is emphasized by this comparison, which also shows how well the attention mechanism works to optimize the fusion process and enhance overall classification performance.in addition to that, The classification report offers a thorough assessment of our suggested model's performance on a 5-way, 4-way, and 3-way multiclass classification problem, as indicated in Tables 4, 5, and 6. This



(a)



(b)



(c)

Figure 6 visually compared in (a) precision, (b) recall, and (c) F1-score of the three models for AD, CN, EMCI, LMCI, and MCI

assignment entails classifying a given instance into discrete classes, each corresponding to a specific disease stage. In particular, the categorization report evaluates how well the model assigns instances to the right classes and how accurate it is in doing so. We can determine how well the model recognizes and differentiates between various illness stages by looking at parameters like precision, recall, F1-score, and overall accuracy.

Table 4 classification report of proposed model for 5-way multiclass

Classes	Precision	Recall	F1-Score	Support
AD	1	1	1	64
CN	.98	.98	.98	52
EMCI	1	.98	.99	57
LMCI	1	1	1	62
MCI	.98	1	.99	53

Table 5 classification report of proposed model for 4-way multiclass

Classes	Precision	Recall	F1-Score	Support
AD	1	1	1	56
CN	.98	.96	.97	67
EMCI	.98	.98	.98	59
MCI	.96	1	.98	52

Table 6 classification report of proposed model for 3-way multiclass

Classes	Precision	Recall	F1-Score	Support
AD	1	.98	.99	50
CN	1	1	1	62
MCI	.98	1	.99	63

ROC curves for four distinct models—DenseNet121, ResNet50, Custom CNN, and a suggested method are shown in Figure 7. These curves, which show distinct phases of Alzheimer's disease and associated cognitive impairments, are drawn for different combinations of classes (0–5).AD, MCI, CN, LMCI, and EMCI are represented by classes 0, 1, 2, 3, 4, and 5, respectively. For a particular classifier, each ROC curve displays the trade-off between the True Positive Rate and False Positive Rate. The classifier's total performance is gauged by its

AUC, where larger values denote superior performance. Across all classes, the suggested method had the highest average AUC (0.99.2), followed by Custom CNN (0.98), DenseNet121 (0.96), and ResNet50 (0.95).

Furthermore, according to the AUC metric, our proposed model demonstrated superior performance in distinguishing among all disease classes. The model achieved a perfect accuracy of 1.0 for classifying AD cases and exceptionally high accuracies of 0.98 for MCI, 0.99 for CN, 0.99 for EMCI, and 1.0 for LMCI. Figure 8 also clearly demonstrates the exceptional performance of our proposed technique. When applied to both 4-way and 3-way classification tasks, the technique consistently achieved a remarkable mean AUC of nearly 99%. This indicates an extremely high level of accuracy in distinguishing between the different classes. For the 4-way classification, the AD, MCI, CN, and EMCI classes exhibited AUC scores of 1.00, 0.97, 0.98, and 0.99, respectively. While for the 3-way classification, the AD, MCI, and CN classes exhibited AUC scores of 0.99, 1.00, and 0.99, respectively. These findings imply that the suggested method is the best model in this dataset for categorizing Alzheimer's disease and associated cognitive deficits.

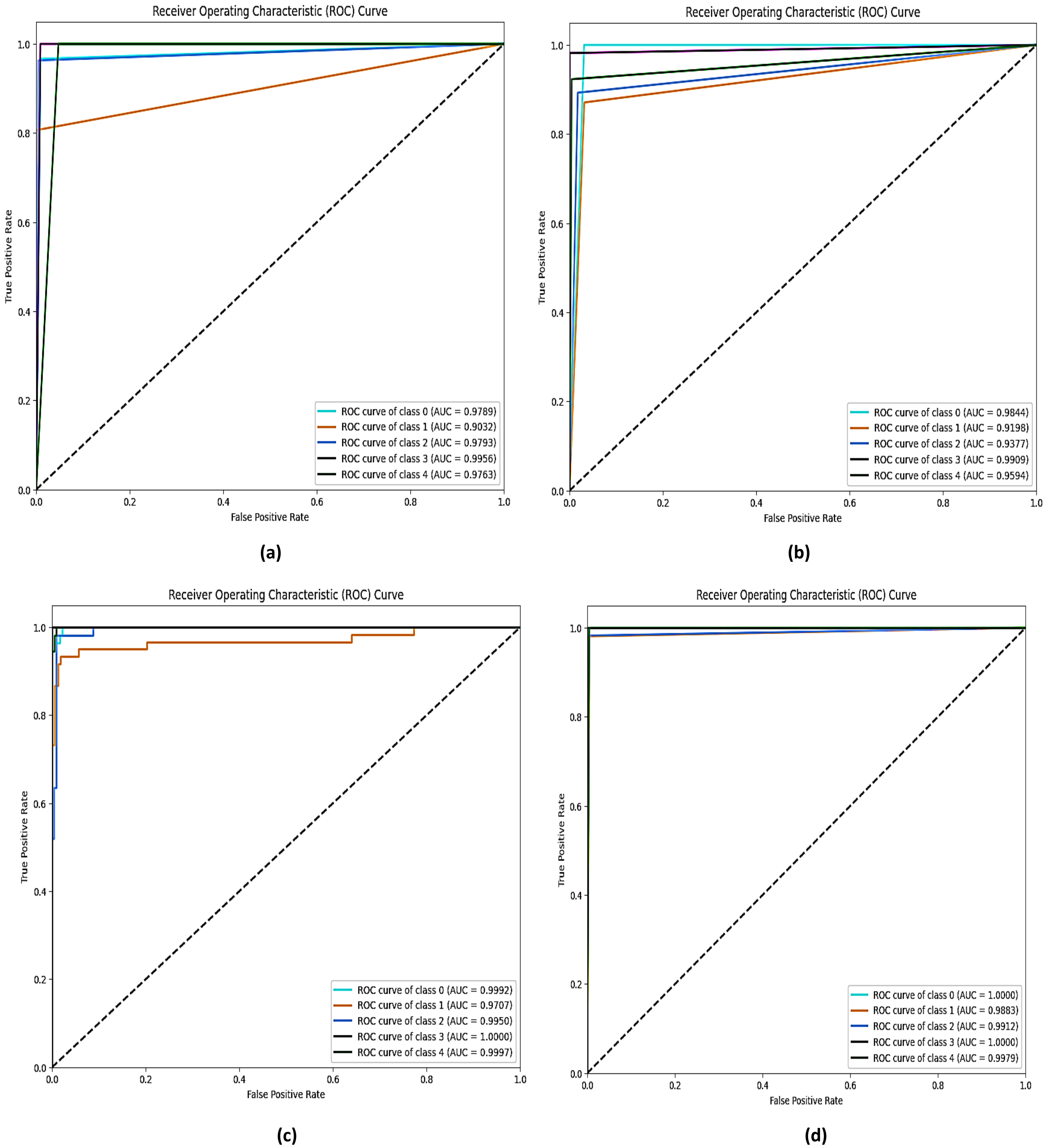
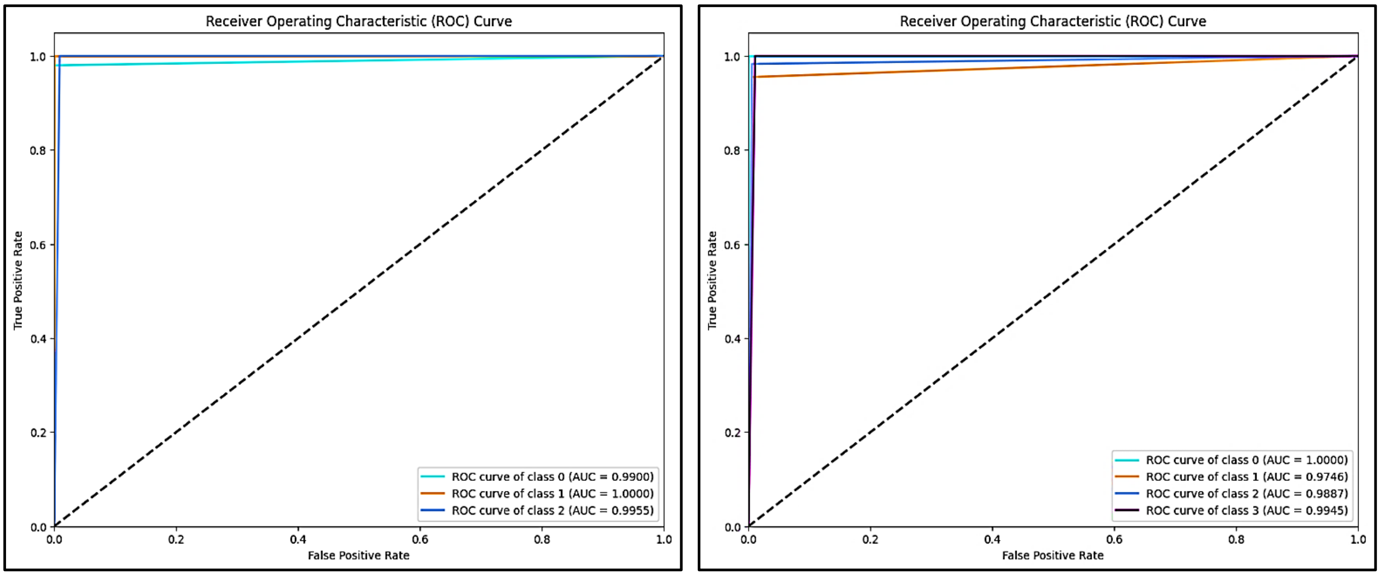


Figure7. ROC curve and AUC value across five categories of the following: (a) DenseNet121, (b) Resnet50v2, (c) custom CNN, and (d) proposed method.



(a) **(b)**

Figure 8 ROC curve and AUC value of the following: (a) proposed method across three categories; and (b) proposed method across four categories.

5.1 Confusion Matrix

Figure 9 shows confusion matrices for the various models used to diagnose Alzheimer's disease. Each model's accuracy in identifying the various stages of Alzheimer's disease (AD, CN, EMCI, LMCI, and MCI) varies. Each matrix's diagonal elements show accurate predictions, while the off-diagonal elements show incorrect classifications[41].

There are noticeable differences in performance when comparing several models for Alzheimer's disease diagnosis, as seen by their confusion matrices. In all areas, DenseNet121 exhibits a high degree of classification accuracy, especially when it comes to recognizing patients with AD and CN. ResNet50 continues to perform admirably, slightly outperforming DenseNet121 in correctly classifying AD cases but exhibiting somewhat greater misclassification in other categories. Although the custom CNN model performs marginally worse overall than the first two, it exhibits good accuracy. Higher misclassification rates result from the model combination lacking attention, which struggles more notably to differentiate CN from EMCI and LMCI. On the other hand, the suggested method performs better in accurately detecting AD, CN, EMCI, LMCI, and MCI instances and attains the best accuracy, hence lowering misclassification. This demonstrates the robustness and enhanced diagnostic potential of the suggested method in comparison to the other models.

6. Discussion

When several deep neural networks are concatenated without a way to synchronize their outputs, several problems may occur. First, certain models may extract comparable features, resulting in feature redundancy and noise in the composite representation. The second issue features misalignment, which can occur when features taken by various models do not line up properly, making it more difficult for the combined model to identify significant patterns. Finally, adding more than one model might increase the model's complexity and increase its susceptibility to overfitting. Soft attention, which functions as a weighting mechanism that gives various input components varying degrees of emphasis, can help to reduce these problems. Several advantages can be obtained by giving the outputs of individual model's careful consideration before concatenation. The quality of the combined representation is improved, and noise is decreased by concentrating on the most pertinent features

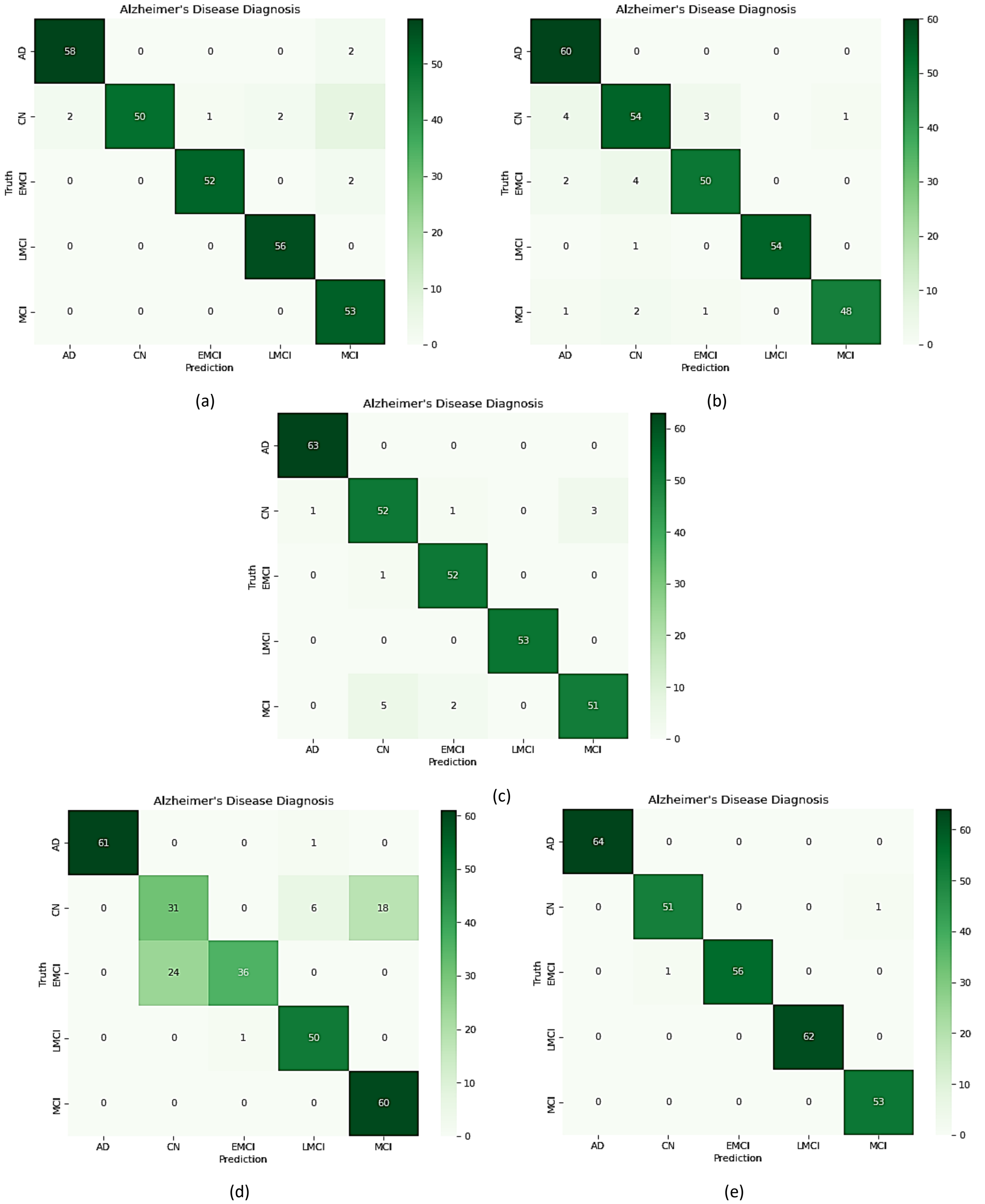


Figure 9 confusion Matrix across five categories of the following: (a) DenseNet121, (b) Resnet50v2, (c) custom CNN, (d) combination without attention, and (e) proposed method.

Table 7 Classification performance comparison

Authors	Biomarker	Database	Methodology	Classification	Accuracy
Shao et al. (2020) [45]	MRI and FDG-PET	ADNI	Hypergraph based regularization	Binary classification AD vs. MCI	92.51%
Tian Zhu et al. (2020) [46]	MRI	ADNI and MIRIAD	DAG-CNN	Binary classification AD vs. NC	91.57%
Fu'adah et al. (2021) [47]	MRI	ADNI	AlexNet	4-way AD/CN/EMCI/LMCI	95%
Lao et al. (2022) [48]	sMRI + PET	ADNI	3D discrete wavelet transform and 3D moment invariant features	Binary classification sMCIs and pMCIs	87.8%
Ko et al. (2022) [49]	sMRI + Clinical + SNP	ADNI	CNN	Binary classification sMCIs and pMCIs	71.6%
Alqahtani et al. (2023) [50]	MRI	ADNI	Deep Belief Networks (DBN) with Optimization Algorithm (MOA)	Binary classification AD vs. NC	97.45 %
Pasnoori et al. (2024) [51]	MRI	Kaggle	Adaptive multi-thresholding algorithm based classification	4-way AD/CN /EMCI/LMCI	96 %
A Rizwan et al. (2024) [27]	MRI	Kaggle+ ADNI	Bi-Vision Transformer	4-way AD/CN /EMCI/LMCI	96.38%
Assmi et al. (2024) [52]	MRI	Kaggle	VGG-19 VGG-16 Inception-V3 Xception ResNet-50 DenseNet169	4-way AD/CN/EMCI/LMCI	92.86% 92.83% 91.04% 90.57% 85.99% 88.64%
Wang et al. (2025) [29]	sMRI + SNP + RGV	ADNI	CNN+ adaptive neural network	Binary classification sMCIs and pMCIs	94.37 %
Proposed	**MRI**	**ADNI**	**Combined ResNet-50v2 DenseNet121, and CNN with soft attention**	**3-way AD/CN/MCI** **4-way AD/CN/EMCI/LMCI** **5-way AD/CN/MCI/EMCI/ LMCI**	**99.43 %** **98.29 %** **99.31%**

that each model has extracted. Second, by weighing the features, it can align the feature spaces of various models, which increases the combined model's learning efficiency. Finally, by enhancing the gradient flow across the network, it can make training easier.

Like many studies, we compare our results with the classification results for each approach in Table 7 to make sure that our method has the best classification performance based on trials in recent studies. These studies differ from one another in terms of the dataset, the number of participants, and the separation of training and testing samples. Thus, there may be some differences in the definition and selection of AD, CN, LMCI, EMCI, and MCI. Due to irregular dataset partitioning and varying subject numbers, the results in Table 7 may not be entirely comparable; nonetheless, we may still generally compare our approach with these cutting-edge techniques to confirm the efficacy of our suggested approach. As indicated in Table 7, the suggested approach in our study performed better in predicting AD across a different set of classification tasks.

Based on all the above, we can say that the proposed method has many advantages, the most important of which are:

1. 1. The usefulness of the suggested method in categorizing Alzheimer's disease is demonstrated by its remarkable accuracy rates, which are 99.31% for five stages, 98.29% for four stages, and 99.43% for three stages. 2. The classification performance is much improved by the integration of multimodal characteristics with the attention mechanism, proving its superiority over conventional single-modality techniques.
2. The significance of attention-based integration is demonstrated by the significant performance improvement that occurs when the attention mechanism is incorporated prior to the feature fusion process.
3. To ensure robustness and reliability, the approach was extensively verified using a variety of performance indicators, including accuracy, recall, Precision, ROC curve, and confusion matrix.

7. Conclusions

This work presents a novel attention-based multimodal feature fusion module for Alzheimer's disease classification. We determined the best techniques for precise categorization by utilizing the ADNI dataset and contrasting different pre-trained models. Before feature fusion, the attention mechanism was incorporated, which greatly enhanced performance and produced high accuracy throughout several classification stages. Our findings show that this approach has promise for Alzheimer's disease early diagnosis and therapy planning. Future research will concentrate on examining the applicability of our methodology in actual clinical settings and further validating it using a variety of datasets.

8. References

1. Xu, H., et al. A Multi-scale Attention-based Convolutional Network for Identification of Alzheimer's Disease based on Hippocampal Subfields. in 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2022. IEEE.
2. Kaur, C., et al. Utilizing the Random Forest Algorithm to Enhance Alzheimer’s disease Diagnosis. in 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS). 2023. IEEE.
3. Wu, Y., et al. Early diagnosis of Alzheimer's disease based on VGG cascade model. in 2022 16th ICME International Conference on Complex Medical Engineering (CME). 2022. IEEE.
4. Zaabi, M., et al. Alzheimer's disease detection using convolutional neural networks and transfer learning based methods. in 2020 17th International Multi-Conference on Systems, Signals & Devices (SSD). 2020. IEEE.
5. El-Assy, A., et al., A novel CNN architecture for accurate early detection and classification of Alzheimer’s disease using MRI data. Scientific Reports, 2024. 14(1): p. 3463.
6. Joshi, R., P. Negi, and T. Poongodi. Multilabel Classifier Using DenseNet-169 for Alzheimer's disease. in 2023 4th International Conference on Intelligent Engineering and Management (ICIEM). 2023. IEEE.

7. Lu, P., et al. An Alzheimer's disease classification method based on ConvNeXt. in 2023 3rd International Symposium on Computer Technology and Information Science (ISCTIS). 2023. IEEE.
8. LeCun, Y., Y. Bengio, and G. Hinton, Deep learning. nature, 2015. 521(7553): p. 436-444.
9. Khan, R., et al., A transfer learning approach for multiclass classification of Alzheimer's disease using MRI images. Frontiers in Neuroscience, 2023. 16: p. 1050777.
10. Al-Shoukry, S., T.H. Rassem, and N.M. Makbol, Alzheimer's diseases detection by using deep learning algorithms: a mini-review. IEEE Access, 2020. 8: p. 77131-77141.
11. Shamrat, F.J.M., et al., AlzheimerNet: An effective deep learning based proposition for alzheimer's disease stages classification from functional brain changes in magnetic resonance images. IEEE Access, 2023. 11: p. 16376-16395.
12. Tanveer, M., et al., Classification of Alzheimer's disease using ensemble of deep neural networks trained through transfer learning. IEEE Journal of Biomedical and Health Informatics, 2021. 26(4): p. 1453-1463.
13. Hajamohideen, F., et al., Four-way classification of Alzheimer's disease using deep Siamese convolutional neural network with triplet-loss function. Brain Informatics, 2023. 10(1): p. 5.
14. Divya, R., R. Shantha Selva Kumari, and A.s.D.N. Initiative, Genetic algorithm with logistic regression feature selection for Alzheimer's disease classification. Neural Computing and Applications, 2021. 33(14): p. 8435-8444.
15. Hazarika, R.A., et al., An improved LeNet-deep neural network model for Alzheimer's disease classification using brain magnetic resonance images. IEEE Access, 2021. 9: p. 161194-161207.
16. Sharma, R., et al., FDN-ADNet: Fuzzy LS-TWSVM based deep learning network for prognosis of the Alzheimer's disease using the sagittal plane of MRI scans. Applied Soft Computing, 2022. 115: p. 108099.
17. Rahul, K., Rider Optimization Algorithm (ROA): An optimization solution for engineering problem. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 2021. 12(12): p. 3197-3201.
18. Borkar, P., et al., Deep learning and image processing-based early detection of Alzheimer disease in cognitively normal individuals. Soft Computing, 2023: p. 1-23.
19. Fareed, M.M.S., et al., ADD-Net: an effective deep learning model for early detection of Alzheimer disease in MRI scans. IEEE Access, 2022. 10: p. 96930-96951.
20. Sharma, S., et al., A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer Disease using MRI scans. Measurement: Sensors, 2022. 24: p. 100506.
21. Sethi, M., et al., A CAD system for Alzheimer's disease classification using neuroimaging MRI 2D slices. Computational and Mathematical Methods in Medicine, 2022. 2022(1): p. 8680737.
22. Liu, Z., et al., Diagnosis of Alzheimer's disease via an attention-based multi-scale convolutional neural network. Knowledge-Based Systems, 2022. 238: p. 107942.
23. Das, R. and S. Kalita. Classification of Alzheimer's Disease Stages Through Volumetric Analysis of MRI Data. in 2022 IEEE Calcutta Conference (CALCON). 2022. IEEE.
24. Phankokkruad, M. and S. Wacharawichanant. Stages of Progression Classification of Alzheimer's Disease Using Deep Transfer Learning Models with Over-Sampling. in 2022 International Conference on Data and Software Engineering (ICoDSE). 2022. IEEE.
25. Alp, S., et al., Joint transformer architecture in brain 3D MRI classification: its application in Alzheimer's disease classification. Scientific Reports, 2024. 14(1): p. 8996.
26. Shaffi, N., et al., Performance evaluation of deep, shallow and ensemble machine learning methods for the automated classification of Alzheimer's disease. International Journal of Neural Systems, 2024. 2450029.
27. Shah, S.M.A.H., et al., Computer-aided diagnosis of Alzheimer's disease and neurocognitive disorders with multimodal Bi-Vision Transformer (BiViT). Pattern Analysis and Applications, 2024. 27(3): p. 76.
28. Awarayi, N.S., et al., A bilateral filtering-based image enhancement for Alzheimer disease classification using CNN. Plos one, 2024. 19(4): p. e0302358.
29. Xi, Y., et al., Predicting conversion of Alzheimer's disease based on multi-modal fusion of neuroimaging and genetic data. Complex & Intelligent Systems, 2025. 11(1): p. 1-20.
30. Ramani, R., et al., Integrated Multi-Modal 3D-CNN and RNN Approach with Transfer Learning for Early Detection of Alzheimer's Disease. Iranian Journal of Science and Technology, Transactions of Electrical Engineering, 2024: p. 1-25.
31. ADNI. Available from: https://adni.loni.usc.edu/.
32. Kauderer-Abrams, E., Quantifying translation-invariance in convolutional neural networks. arXiv preprint arXiv:1801.01450, 2017.
33. De, A. and A.S. Chowdhury, DTI based Alzheimer's disease classification with rank modulated fusion of CNNs and random forest. Expert Systems with Applications, 2021. 169: p. 114338.
34. Shorten, C. and T.M. Khoshgoftaar, A survey on image data augmentation for deep learning. Journal of big data, 2019. 6(1): p. 1-48.
35. He, H., et al. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. in 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence). 2008. Ieee.
36. Deepa, S., J.L. Zeema, and S. Gokila, Exploratory Architectures Analysis of Various Pre-trained Image Classification Models for Deep Learning. Journal of Advances in Information Technology, 2024. 15(1).

37. He, K., et al. Identity mappings in deep residual networks. in Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14. 2016. Springer.
38. Talukder, M.A., et al., An efficient deep learning model to categorize brain tumor using reconstruction and fine-tuning. Expert systems with applications, 2023. 230: p. 120534.
39. Albelwi, S., Deep Architecture based on DenseNet-121 Model for Weather Image Recognition. International Journal of Advanced Computer Science and Applications, 2022. 13 DOI: 10.14569/IJACSA.2022.0131065.
40. Thakur, M., et al., Soft attention based DenseNet model for Parkinson's disease classification using SPECT images. Frontiers in Aging Neuroscience, 2022. 14: p. 908143.
41. Japkowicz, N. and M. Shah, Performance Evaluation in Machine Learning, in Machine Learning in Radiation Oncology: Theory and Applications, I. El Naqa, R. Li, and M.J. Murphy, Editors. 2015, Springer International Publishing: Cham. p. 41-56.
42. Chicco, D., N. Tötsch, and G. Jurman, The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. BioData Mining, 2021. 14(1): p. 13 DOI: 10.1186/s13040-021-00244-z.
43. Zhang, X., et al., The use of ROC and AUC in the validation of objective image fusion evaluation metrics. Signal processing, 2015. 115: p. 38-48.
44. Google Colab. Available from: https://colab.research.google.com/.
45. Shao, W., et al., Hypergraph based multi-task feature selection for multimodal classification of Alzheimer's disease. Computerized Medical Imaging and Graphics, 2020. 80: p. 101663.
46. Zhu, T., et al., Anatomical landmarks and DAG network learning for alzheimer's disease diagnosis. IEEE Access, 2020. 8: p. 206063-206073.
47. Fu'adah, Y.N., et al., Automated Classification of Alzheimer's Disease Based on MRI Image Processing using Convolutional Neural Network (CNN) with AlexNet Architecture. Journal of Physics: Conference Series, 2021. 1844(1): p. 012020 DOI: 10.1088/1742-6596/1844/1/012020.
48. Lao, H. and X. Zhang, Diagnose Alzheimer's disease by combining 3D discrete wavelet transform and 3D moment invariants. IET Image Processing, 2022. 16(14): p. 3948-3964.
49. Ko, W., et al., A deep generative–discriminative learning for multimodal representation in imaging genetics. IEEE Transactions on Medical Imaging, 2022. 41(9): p. 2348-2359.
50. Alqahtani, N., et al., Deep belief networks (DBN) with IoT-based alzheimer's disease detection and classification. Applied Sciences, 2023. 13(13): p. 7833.
51. Pasnoori, N., T. Flores-Garcia, and B.D. Barkana, Histogram-based features track Alzheimer's progression in brain MRI. Scientific Reports, 2024. 14(1): p. 257.
52. Assmi, A., et al., Alzheimer's disease classification: a comprehensive study. Multimedia Tools and Applications, 2024: p. 1-24.