XDementNET: An Explainable Attention Based Deep Convolutional Network to Detect Alzheimer Progression from MRI data

Soyabul Islam Lincoln
Dept. of Electronics and Communication Engineering
Khulna University of Engineering & Technology
KUET Road, Fulbarigate, Khulna-9203.
islam1709028@stud.kuet.ac.bd

Mirza Mohd Shahriar Maswood
Dept. of Electronics and Communication Engineering
Khulna University of Engineering & Technology
KUET Road, Fulbarigate, Khulna-9203.
shahriar@ece.kuet.ac.bd

Abstract

A common neurodegenerative disease, Alzheimer's disease requires a precise diagnosis and efficient treatment, particularly in light of escalating healthcare expenses and the expanding use of artificial intelligence in medical diagnostics. Many recent studies shows that the combination of brain Magnetic Resonance Imaging (MRI) and deep neural networks have achieved promising results for diagnosing AD. Using deep convolutional neural networks, this paper introduces a novel deep learning architecture that incorporates multiresidual blocks, specialized spatial attention blocks, grouped query attention, and multi-head attention. The study assessed the model's performance on four publicly accessible datasets and concentrated on identifying binary and multiclass issues across various categories. This paper also takes into account of the explainability of AD's progression and compared with state-of-the-art methods namely Gradient Class Activation Mapping (GradCAM), Score-CAM, Faster Score-CAM, and XGRADCAM. Our methodology consistently outperforms current approaches, achieving 99.66% accuracy in 4-class classification, 99.63% in 3-class classification, and 100% in binary classification using Kaggle datasets. For Open Access Series of Imaging Studies (OASIS) datasets the accuracies are 99.92%, 99.90%, and 99.95% respectively. The Alzheimer's Disease Neuroimaging Initiative-1 (ADNI-1) dataset was used for experiments in three planes (axial, sagittal, and coronal) and a combination of all planes. The study achieved accuracies of 99.08% for axis, 99.85% for sagittal, 99.5% for coronal, and 99.17% for all axis, and 97.79% and 8.60% respectively for ADNI-2. The network's ability to retrieve important information from MRI images is demonstrated by its excellent accuracy in categorizing AD stages.

1 Introduction

Throughout life, the human brain experiences intricate, age-related changes. Throughout development, these alterations exhibit non-linear and region-specific patterns. Cognitive function often deteriorates with age, resulting in widespread brain atrophy Cole and Franke (2017). As humans age, brain tissue and nerve cells die gradually by pairs of TAU proteins that stabilize the microtubules, leading to Alzheimer's disease. Some symptoms are memory loss, wording, confounding, and disorientation. In 2018, there were almost 50 million victims worldwide, extending the number to 131 million by 2050, and the semi-economic cost will be 9.12 trillion dollars (Scheltens et al., 2021). Although AD typically affects people 65 and older, the number of cases with early onset suggests that it cannot be purely categorized as an old age disease (Association, 2019). Unfortunately, AD has no known cure, and the medicines that are now available merely slow down the disease's course (Neugroschl and Wang, 2011). Although AD can develop in some persons with mild cognitive impairment (MCI), not all MCI patients go on to acquire AD. The diagnosis of brain disorders relies heavily on a number of neuroimaging methods, including Computed Tomography (CT), Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI), and Functional Magnetic Resonance Imaging (fMRI).The great majority of AD diagnosis methods now in use depend on laborious manual assessment by medical experts. In order to reduce medical

costs, enhance treatment results, and postpone the abnormal degeneration of the brain, researchers are exploring early diagnosis for this disorder (Wu et al., 2022). Modern brain imaging techniques like Positron Emission Tomography (PET) and Magnetic Resonance Imaging (MRI) can be used to identify AD early on. Clinically, MRI is widely used for AD-related diagnosis due to its cost-effectiveness and non-invasiveness.

Numerous recent research provide a range of options for AD detection using MRI images by utilizing machine learning techniques, especially deep learning methodologies. The accuracy of identifying various forms of dementia is improved when machine learning (ML) is used to neuroimaging. Certain pre-processing steps must be taken in order to facilitate the use of ML algorithms. These steps include using a classifier method, reducing the dimensionality of features, and extracting and choosing features. Each of these stages is essential to the ML classification process (Ieracitano et al., 2020). Convolutional Neural Networks (CNNs), which can automatically extract a variety of relevant semantic information from pictures, have been progressively encompassed to MRI diagnostic tasks due to the quick advancement of computing power and computer vision. MRI is regarded the gold standard, due to its limited ionizing radiation, good tissue contrast, and great spatial resolution (Moser et al., 2009). Developing a trustworthy computer-aided diagnostic system that can analyze MRI images and detect AD is essential. As such, any AI algorithm used to detect dementia in an MRI scan must be able to classify the stage of the disease so the person can get the proper treatment. Furthermore, many AI models, such as neural networks, are 'black boxes', which means the researcher is unsure exactly how the model makes the classification based on input (Ebrahimi et al., 2021; Salehi et al., 2020). In the case of an MRI-based diagnosis, the developer is unable to show what regions of the scan the model is using to make its classification. This can cause problems in the diagnosis process, as the medical professional is unable to verify the diagnosis or see the logic behind the AI's classification. To solve this problem, the discipline of Explainable Artificial Intelligence (XAI) was created, which creates new kinds of interpretable models or provides explanations for the predictions made by black-box systems. In the imaging field, creating a unique heatmap for every input image and highlighting the significance of each pixel in the ultimate classification choice is a popular method. Such a technique can be an efficient tool that can be effortlessly incorporated into prospective computer-aided diagnosis software to offer human-intuitive justifications for the classifier's choices Böhle et al. (2019).

Despite the severity of these obstacles, scientists and medical professionals are making every effort to overcome them in the context of classifying neurodegenerative diseases. The main focus of these initiatives is the use of AI methods that help produce precise diagnostic results. This research has focused on those those problem and attempted to work through it. The major contributions of this research are as follows:

- We present a novel deep learning architecture that employs an advanced convolutional neural Network (CNN), with multiresidual block, custom spatial attention block, grouped query attention and Multihead attention.
- Four datasets were used in the development and testing of a generalized model, which led to lower computing costs and parameters. The model performed exceptionally well in the diagnosis of AD in spite of these optimizations.
- Rather than working on a single plane (axial), our model was trained and evaluated on other two planes i.e. sagittal, coronal, and also combining all of the planes.
- We utilized five state of the art XAI techniques such as, GradCAM, ScoreCAM, Faster ScoreCAM, and XGradCAM to provide interpretable explanation for predictions generated by our proposed approach. Because it allowed us to confirm the significance of particular features, compare and contrast the data acquired from other methodologies, and highlight the clinical relevance of our model, the interpretability was compared with other current approaches.
- We have trained and evaluated our model in different class scenarios such as 5 stages of AD, 4 stages, 3 stages, and binary classification of AD.
- Our results show that, in spite of the dataset's class imbalance, our model performed exceptionally well thanks to efficient feature engineering and modeling techniques. This could have important ramifications for enhancing ML applications and promoting their use in the clinical setting.

The remaining paper is organized as follows: section 2 presents a comprehensive analysis of the current literature on the classification of AD. section 3 discusses the detail overview of dataset, their pre-processing techniques and brief overview of proposed architecture with its parameters and implementation details followed by the explanation of the result in section 4 with explainability analysis. Finally section 5 gives the conclusion with future direction.

2 Related Works

The approach of using machine learning to AD progression modeling is intricate. Most AD studies relies on MRI analysis. Kavitha et al. (2022) incorporated voting classesfier with different ML models on OASIS dataset namely:

Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost). To identify most relevant features from the dataset, some feature selection (FS) algorithms was used. Using the same dataset, another work was done by comparing different ML models, and it was observed that the SVM model gains highest accuracy amongst other models (Bari Antor et al., 2021). Another work incorporated PET data along with MRI scan data to predict the progression from MCI to AD using state-of-the-art ML models (SVMRF, Gradient Boosting (GB)) (Beltran et al., 2020). Rallabandi et al. (2020) implemented nonlinear SVM on whole brain MRI scan to classify the progression of brain stages of impairment from cognitive normal (CN) to AD. Similarly, in another work by Shahbaz et al. (2019), experiment was conducted on ADNI datasets using different ML models like Naive Bayes(NB), DT, rule reduction, and generalized linear model. Almohimeed et al. (2023) suggested a unique multi-level stacking approach to predict various AD phases by combining heterogeneous ML models. Six base model was used in this study namely-RF, DT, Logistic Regression (LR), k-nearest Neighbors(KNN), and NB. Model interpretability was also introduced to ensure the efficiency, effectiveness of proposed model using Explainable Artificial intelligence(XAI). Similarly, data-level fusion using Alzheimer's Disease Research Center (ADRC) clinical data, brain Magnetic Resonance Imaging (MRI) segmentation data, and psychological assessments data has been used in the work of Jahan et al. (2023a). Different ML models was used along with SHAPELY for model interpretation. Fusion of different modalities data could provide a better view on patient's situation and could be able to enhance the accuracy of progression model. Zhang et al. (2012) suggested a generic technique known as multimodal multitask (M3T) learning that combines multimodal data (such as MRI, PET, and CSF) to predict various medical scores (such as the MMSE, ADAS, and diagnostic feature) at the same time. They used baseline data from a limited number of modalities and SVMs. El-Sappagh et al. (2021) developed a two-layered explainable machine learning model for the categorization of AD. Data from eleven different modalities—genetics, medical history, MRI, Positron Emission Tomography (PET), neuropsychological battery, cognitive scores, etc.—were combined in this multimodal method. Here, the Random Forest (RF) classifier was employed in the first layer for multiclass classification, and the SHAP framework was utilized to interpret the findings. Binary categorization, which likely classified MCI to AD, occurred in the second layer. The majority of AD research only looks at a small number of characteristics, which may not be enough to fully comprehend this complicated illness (Ding et al., 2018). Though these works show very good performance , but these technique solely are dependent on good extracted features, which is time consuming and needs experts' opinions. Deep learning (DL) can improve this challenging modeling task's performance (Zhang and Yang, 2021).

Most of the state-of-the-art DL models either utilize pre-training or non-pretraining methods. Some literatures have used transfer learning techniques on models including VGG16, ResNet, AlexNet, GoogleNet, and etc. Jain et al. (2019) proposed a CNN transfer learning architecture (VGG-16) for AD diagnosis. The results of their experiment are based on a three-way categorization in the ADNI database. In another work, Lee et al. (2019a) introduced a deep CNN data permutation strategy that uses resting-state MRI to classify AD. To make best use of AlexNet's advantages, they suggested slice selection. Their data permutation approach enhanced the overall classification accuracies for AD classification, according to their experimental results. In another work, rather than using conventional thresholding, AbdulAzeem et al. (2021) utilized adaptive thresholding that changes dynamically over the image. It helps to adapt to changes in lighting conditions of images. Simple CNN architecture has been used for predicting AD progression using ADNI dataset and several experiments was done by changing parameters to get the best result. Shamrat et al. (2023) worked with some basline pre-trained architectures such as ResNet50, MobileNetV2, VGG16, AlexNet and Inception-V3 by fine-tuning the parameters and finally proposed a modified InceptionV3 as the baseline outperforms other architectures. Two different CNN architecture was proposed by El-Assy et al. (2024)consisting different filters for different feature extraction, which are then concatenated before the decision layer. Thus, both knowledge of task specific features is enabled by complimenting the models to each other. Adaptive Synthetic Sampling Approach (ADASYN) has been utilized for solving the data imbalance problem. Arafa et al. (2024) proposed a simple CNN architecture along with another method that uses a pre-trained VGG16 architecture and compared the output of both on different parameters by applying some pre-processing on dataset and data augmentation to mitigate the class imbalances. Similarly, in another study, a joint learning technique that combines both deep learning and machine learning approach was proposed by Abuhmed et al. (2021). Two hybrid architecture was introduced that is able to predict the AD progression even after 2.5 years later from the multimodal time series data. The two methods are: deep feature-based learning which utilizes multivariate BiLSTM architecture for feature learning following the ML models for prediction and multitask regression-based learning that first learns the seven regression tasks and then ML models for prediction. Lee et al. (2019b) employed a multimodal recurrent neural network (RNN) model for predicting the progression of AD from the stage of MCI. In this method, the authors combined cross-sectional neuroimaging data, demographic data, and the participants' longitudinal cerebrospinal spinal fluid (CSF) and cognitive performance biomarkers. In another study, multiclass classification was performed on volumetric 18F-FDG PET images collected from ADNI dataset (De Santi et al., 2023). Saliency Map (SM) and Layerwise Relevance Propagation (LRP) are two distinct post hoc explanation strategies that were used to test the performance of a 3D CNN architecture.Similar to this, a modified Resnet18 deep learning architecture is trained concurrently on the two datasets using a unique heuristic early feature fusion approach

that concatenates PET and MRI images. Ansingkar et al. (2022) adopted a hybrid equilibrium approach to improve a capsule encoder network for AD diagnosis. A hybrid CNN model, for feature extraction, followed by a KNN with Bayesian optimization for thewas incorporated in another study (Lahmiri, 2023). The validation dataset's sample size is constrained, though. In addition to the multi-layer CNN structure, autoencoders were developed in this study to increase accuracy by utilizing fine-grained abstract information.

To increase the efficacy of CNN-based models, enhanced attention-guided multiscale features can be used as the attention mechanism canprovide more focused features on weights and provide proper interpretability. In several literature, different types of attention mechanism such as channel-wise squeeze and excitation module, spatial attention, self attention, and different task specific attention module has been used (Liu et al., 2022; Zhang et al., 2021b,a; Jin et al., 2019). These studies shows the effectiveness of attention modules in sense of model performance and interoperability. By addressing other literatures lackings on attention to high or low-level spatial features, Tripathy et al. (2024) proposed a modified spatial attention guided network with the depth separable convolutional layer. The proposed approach introduces improved spatial attention block to enhance the spatial attention maps by combining multiple feature cues. These maps are used to extract multiscale, spatially directed features. The maps goes through the separable CNN layer with skip connection and multilayer spatial attention feature is combined forming scale invariant features that goes into the classification stage. Most of the proposed attentionmethods work with single spatial scale which are unable to learn a discriminative feature representation of AD patients. So, a multi-scale information integration block has been proposed by Wu et al. (2022) using dilated convolution and soft attention mechanism. It helps to capture better multi-scale features with fewer computations. 3D MRI images consists better multi-sclare features, which was studied also in several research. A powerful interpretable network naming TabNet, which was built upon transformer model, was proposed for AD classification using 3D t1-weighted brain MRI dataset (Park et al., 2023). The dataset was processed to extract 104 brain subregion and 68 cortex region by the help of VUNO Med-deepbrain , which is a deep learning based software to analyze brain parcellation and quantification. Kang et al. (2023) proposed a framework for extracting high dimensional brain ROI's with the help of CNN. The DL based biomarkers has been used by Explainable boosting machines (EBM), a tree-based Generalized Additive Model (GAM), to predict AD progression. As baseline methods to extract biomarkers from brain, Glo-CNN and Loc-CNN has been incorporated. In another study, Dhaygude et al. (2024) proposed a deep 3D convolutional network incorporating multitask learning and attention mechanism. Two auxiliary subtasks such as Clinical Dementia Rating (CDR) scale score regression and Mini-Mental State Examination (MMSE) score regression is added to optimize the accuracy of AD classification. For the purpose of randomizing clinical trials for AD and assessing the effect of reducing allocation bias on trial efficiency, a second CNN-based model with a self-attention mechanism was presented by Wang et al. (2024). Non-image data, including as demographics, cognitive test results, and biomarkers, were added to T1-weighted pictures in order to forecast changes in the patient's health. In the study by Mahim et al. (2024) for AD classification vision transformer (ViT) with the combination of GRU has been incorporated. The ViT extracts crucial features from the images and GRU creates strong correlation between those features giving the architecture a push to higher performance. Different XAI techniques was experimented namely LIME, SHAP and attention maps for clear model interpretation. Ahanger et al. (2024) introduced Alzhinet that uses self-attention mechanism on 3D volumetric MRI data to predict AD stages. The baseline pretrained VGG16 model was utilized rather than more complex structures like ResNet59 and Inception model. The three view of brain namely axial, sagittal and coronal was processed with several steps like skull stripping, denoising, normalization, and registration nad used for classification. Before entering the VGG16 network as a 3-channel, 224x224 pixel 2D picture, each MRI slice of the scan is scaled along the axial plane's z-axis. Adarsh et al. (2024) presented a CNN framework and a within-class-similar discriminative dictionary learning technique that uses anatomical and structural similarities between similar pictures in the training set to minimize misclassification. To verify and enhance the classification's accuracy, a transfer learning procedure and a decision tree mechanism are employed. To create a model that is comprehensible and whose judgments can be relied upon, LIME and CAM are utilized.

3 Materials and Methods

We introduce a novel framework intended to accomplish two important objectives in the categorization of medical images: interpretability and high accuracy. For efficient classification, our system combines advanced CNN techniques, kernel approaches, and explanation algorithms. Each of the framework's several parts addresses a distinct set of difficulties in the categorization and interpretation of medical images. We have carefully processed medical photos before feeding them into the neural network. Data is splitted in three protions: train, validation, and test. First the test data, which is 15% of the total data, is separated. Then, from the remaining data train and validation data has been splitted in 85:15 ratio. Image augmentation is performed only on the training data to mitigate the problem of huge class imbalancing. A specially created CNN that is trained for categorization, sits at the heart of the system. We have included a Group Query Attention layer, a multihead attention layer, and a bespoke Spatial Attention layer to improve the procedure. The network can also extract characteristics at different degrees of detail because of the integration of a

multi-residual structure. The proposed network is trained and evaluated on train and validation data respectively, and later tested on the test dataset in different sets of experimentation. The overall overview of the proposed methodology is shown in Figure 1.

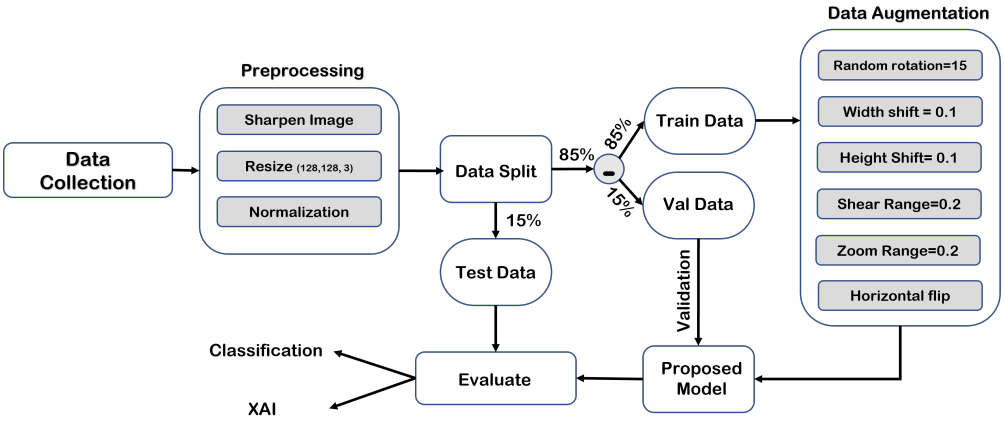


Figure 1: An overview of the proposed method

3.1 Dataset

3.1.1 Kaggle Dataset

This research uses an publicly available dataset to detect the progression of AD from the open-source data science platform Kaggle (Pinamonti, 2021). The dataset comprises a total of 6,400 images, outsourced from different hospitals, websites, and public repositories. The data set consists of four classes, namely moderate demented, mild demented, very mild demented, and non-demented each having a subjects of 2, 28, 70, and 100, respectively. From each subject scan 32 slice images are extracted. Some sample images of four classes of this dataset are shown in Figure 2.

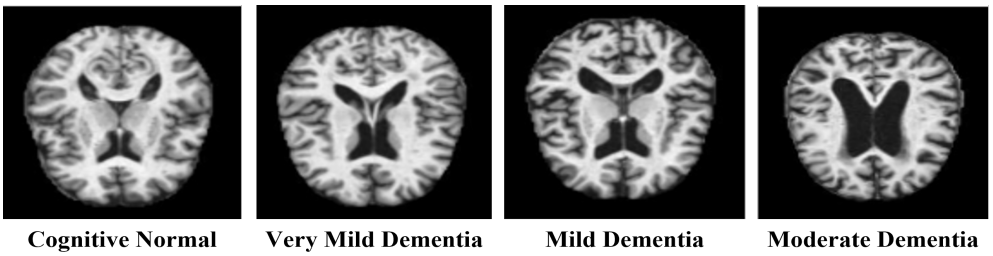


Figure 2: Sample images of kaggle datasets on 4 classes

3.1.2 Open Access Series of Imaging Studies (OASIS)

The OASIS dataset hosted by Neuroimaging Informatics Tools and Resources Clearinghouse-Image Repository (NITRC-IR), provides community support for easy access. In this experiemnt, we used the OASIS-1 dataset. The dataset contains a cross-sectional collection of 416 subject from middle aged person to elderly person (18 - 96 years old). For each subject, 3 or 4 individual T1-weighted MRI scans obtained in single scan. The subjects include both men and women, additionally it was kept in mind that all the subjects are right handed. A clinical diagnosis of very mild to severe AD has been made for 100 of the participants over 60. Furthermore, 20 nondemented participants who were photographed on a follow-up visit within 90 days of their original session make up the reliability data set. From this dataset used in this study, each subject's scans is sliced along sessions are included. The z-axis or plane into 256 pieces, and slices ranging from 100 to 160 are selected from each recordings. Totalling the dataset of 80,000 images.

Table 1: Overview of sample image distribution of each class in datasets

Classes	**Datasets**					
	Kaggle	*OASIS*		*ADNI-1*		*ADNI-2*
			Axial	Sagittal	Coronal	
Cognitive Normal (Non-Demented)	3200	67,222	3880	3880	3880	8,650
Early Mild Cognitive Impairment (Very Mild demented)	2240	13,725	-	-	-	480
Mild Cognitive Impairment (Mild Demented)	896	5002	6320	6320	6320	1155
Late Mild Cognitive Impairment	-	-	-	-	-	144
Alzheimer Disease (Moderate Demented)	64	488	2660	2660	2660	8346

3.1.3 ADNI-1

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a large, collaborative effort established to help researchers track the progression of AD using various biomarkers, including neuroimaging, genetic markers, and cognitive measures. ADNI-1 specifically focused on collecting baseline data and longitudinal follow-ups of participants diagnosed with AD, MCI, and cognitively normal (CN) subjects One of the key imaging techniques used in ADNI-1 is MRI, performed with both 1.5T scanners. This dataset contains recording of 643 subjects. The scans of each subject was divided into three planes namely axial, coronal and sagittal. From each plane among different slices only middle 20 slices were converted into image (Petersen et al., 2010). In the Figure 3 below, shows a brain slice of three plane.

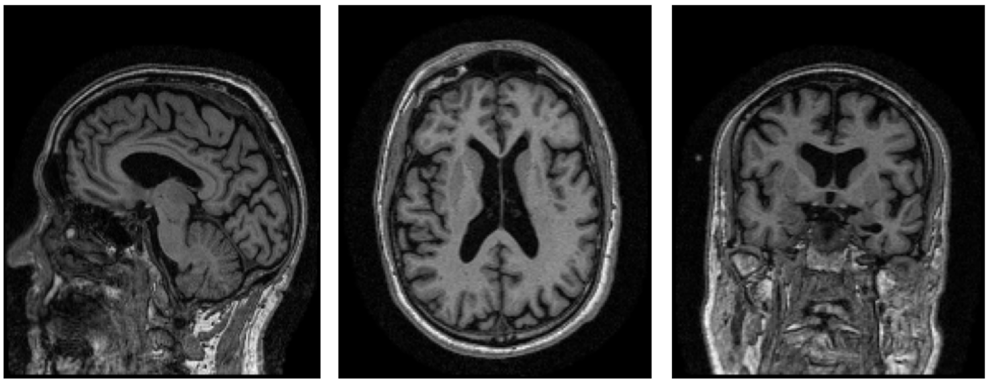


Figure 3: Sliced image of three plane of brain

3.1.4 ADNIGO/ADNI-2

ADNI-2, also known as ADNIGO (ADNI Grand Opportunity), expanded the original ADNI-1 study to further investigate the progression of AD and improve early detection of the disease. ADNI-2, which ran from 2010 to 2016, introduced several advancements in imaging technology and included a wider array of participants, including individuals with early mild cognitive impairment (EMCI) and late MCI (LMCI), to capture an even broader spectrum of the cognitive aging process (Petersen et al., 2010). This dataset contains 5 classes of T1 weighted and T2 weighted processed datasets. Only axial plane has been selected from the scans of each subject. Among the whole slices of the axis, middle 16 slices has been considered.

3.2 Data partition and class distribution

In our experiment, we evaluate four open-access datasets with a variety of class segments. For each study, we ensure the test data remains separate and consistent across all experiments for each dataset. To achieve this, we first split the data into 85% and 15%, reserving the latter for testing. We then further divide the initial 85% into training and validation sets with same 85:15 ratio. By using a consistent random state of 43 during the first stage of splitting, we ensure the train, validation, and test data remains the same for every dataset across all experiments.

Each dataset consists of different classes. We experiment on various extents to evaluate our proposed method. For both the OASIS and Kaggle datasets, four classes are available: Moderate Demented, Non-Demented, Very Mild Demented, and Mild Demented. We experiment with both multiclass and binary classification. In the four-class setup, the classes remain unchanged. For the three-class setup, we combine Mild Demented and Moderate Demented patient data. For binary classification, we merge Mild Demented and Very Mild Demented samples with Moderate Demented samples. In the ADNI-GO dataset, there are five classes: AD, CN, EMCI, MCI, and LMCI. We conduct two multiclass experiments using five classes and three classes. For the three-class setup, we combine EMCI and LMCI data with MCI. In the ADNI-I dataset, we create four datasets using each brain axis and one with all axis. Using this dataset, we conduct only multiclass classification using the existing three classes: AD, MCI, and CN.

3.3 Preprocessing and Data Augmentation

Data preprocessing is a vital step in image classification tasks, ensuring the data is properly prepared before putting inside the model. Our study utilizes multiple datasets, including the ADNI-I and ADNI-GO datasets, which consist of DICOM format brain scans. These scans are 3D images with dimensions of (256×256×256), captured in axial, coronal, and sagittal planes. However, not all slices are relevant for analysis. From the ADNI-I dataset, we selected 20 middle slices, while for the ADNI-GO dataset, 16 middle slices were extracted. These slices were then converted into JPEG format for further processing. Subsequent preprocessing steps were consistent across all datasets. First, the images were converted to RGB format, followed by rescaling and normalize pixel values between 0 and 1. This normalization is essential, as neural networks are sensitive to input data scales, and rescaling helps the model effectively learn patterns. Additionally, all images were resized to a standardized dimension of (128×128x3) to ensure the uniformity across the dataset and improve processing efficiency. We use the sharpening filter to enhance the edges and fine details. The central pixel is emphasized with a value of five, while neighboring pixels are subtracted using negative weights (-1). This results in a stronger contrast at the edges, making them more defined and more sharpened across the overall image. The overall preprocessing stage can be seen in Figure 1.

The data augmentation process plays a critical role in addressing class imbalancing and enhancing the diversity of the dataset, which ensures a more robust model. The augmentation parameters, in this experiment, are carefully selected to promote variability. We used a rotation range of 15 degrees, width and height shift ranges of 0.1 (10% of the image dimensions), a shear range of 0.2, and a zoom range of 0.2 (up to 20%). We also apply random horizontal flips to further diversify the images. To do this, Tensorflow's built-in TrainDatagen function is used. For the multiclass classification task in this experiment, a strategic augmentation strategy was adopted to balance the dataset. In particular, the underrepresented class's sample size is raised to equal that of the second-largest class. The underrepresented class receives one-third as many samples as the second-largest class when there is a very significant difference in class numbers. This targeted approach ensures that the model is trained on a more balanced and representative dataset, thereby improving its ability to learn more generalized patterns across all classes. It is important to note that the augmentation is done after the data split and only on training dataset. The validation data and test data are free from any data augmentaion.

3.4 Proposed Architecture

The proposed architecture combines multiple advanced components to achieve high accuracy and interpretability in alzheimer progression classification. It begins with processed input images through a series of Conv2D layers with varying filter sizes and strides, followed by Max Pooling for feature extraction and dimensionality reduction. There are several convolutional layer block in the structure, consisting of different orientation of layers and normalization to capture the best porssible features from the input. A custom Spatial Attention block enhances focus on critical regions, while a Multi-Residual structure ensures effective feature extraction and learning. The Group Query Attention layer refines feature extraction by modeling inter-feature relationships, aided by Layer Normalization for stability. A Multi-Head Attention layer captures long-range dependencies, improving feature representation. To prevent overfitting, a Dropout layer is included, and Global Average Pooling reduces feature dimensions into compact vectors. Finally, Dense layers process the extracted features and with the last dense layer performs the classification. This design effectively integrates attention mechanisms, normalization, and pooling strategies to handle complex medical imaging tasks with precision and clarity.The proposed architecture and its internal components are illustrated in Figures 5, 4, and 6.

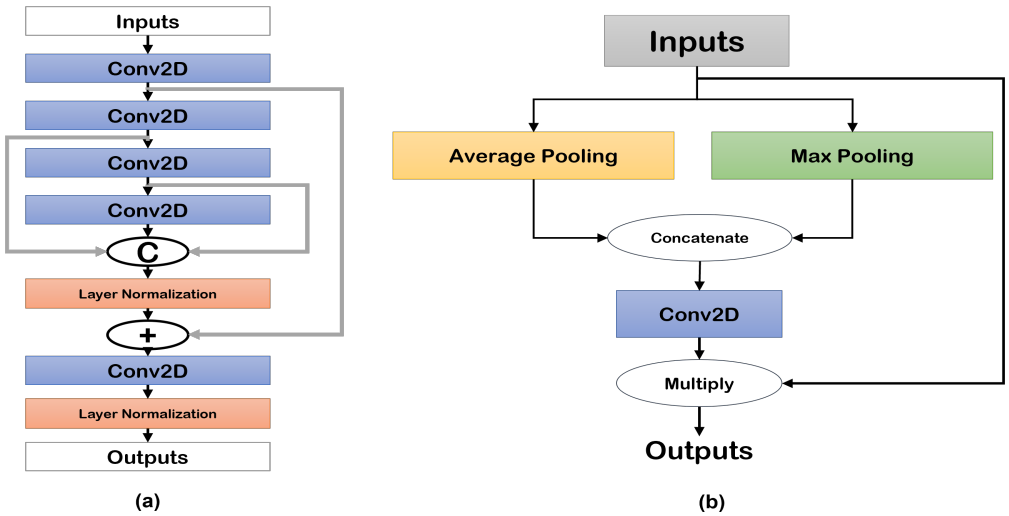


Figure 4: Architecture illustration of - (a) Multi-residual block, and (b) Custom spatial Attention block

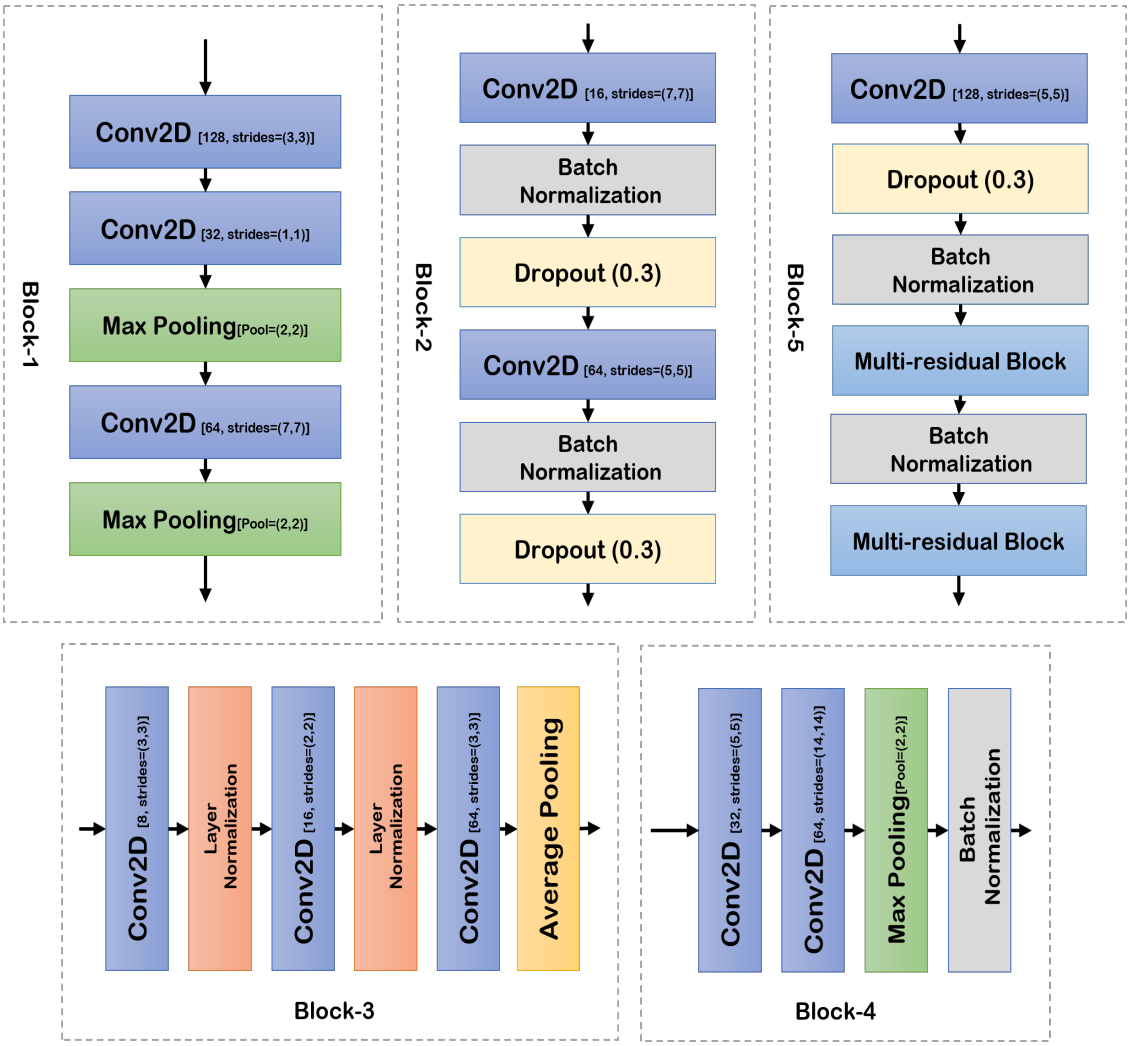


Figure 5: Structure of diffferent convolutional blocks' structure

3.5 Hyperparameters and Optimization Strategy

As we optimize the proposed model, our focus gradullay shifts to hyperparameter tuning, a crucial step in achieving optimal performance. This process involves identifying hyperparameter values that closelyyields the best possible results.We have uniformed the hyperparameters and initialization processes to provide uniform and equitable assessment across various neural network topologies. In all experiments, the same hyperparameters are applied uniformly. Table 2 lists the hyperparameters that we have used in experiments along with their corresponding values.

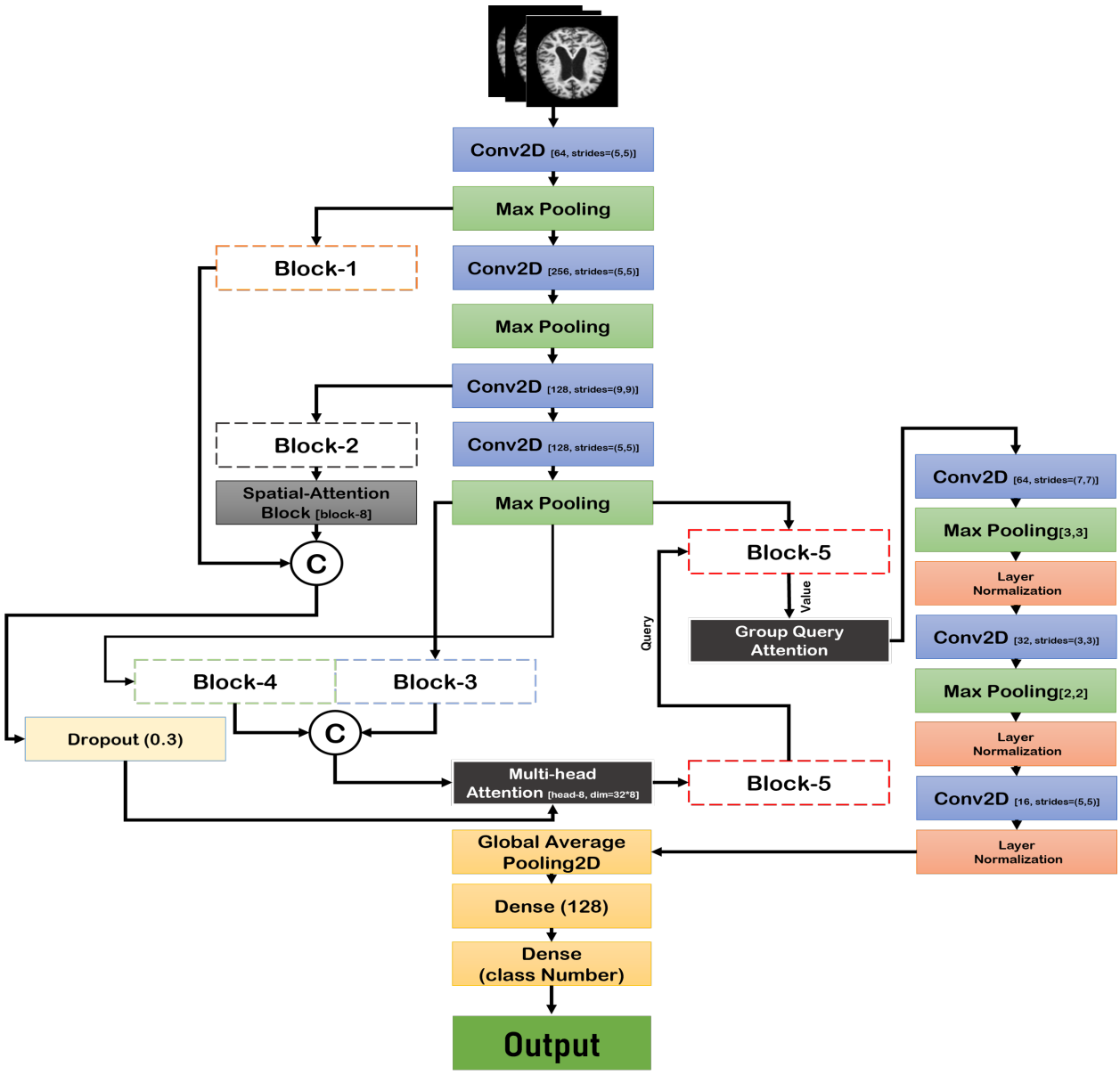


Figure 6: Proposed Model Architecture

Table 2: Hyperparameters of our proposed method and its values

Hyperparameters	*Values*	*Hyperparameters*	*Values*
Epochs	50	Initial Learning Rate	0.0001
Batch Size	8	Factor	0.7
Image Size	128 x 128 x 3	patience	7
Callback	ReduceOnPlateau	min_lr	$1e-6$
Loss	Categorical cross-entropy	Optimizer	Adam

3.6 Evaluation Procedures and Training

The performance of the proposed model is assessed using various metrics, including accuracy, precision, recall, F1-score, and sensitivity. These metrics are derived from the confusion matrix, which provides values such as True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) to reflect the model's performance. Additionally, the Area Under the ROC Curve (AUC-ROC) is calculated using the True Positive Rate (TPR) and False Positive Rate (FPR) to further validate the model. Root Mean Squared Error (RMSE) is also employed as an evaluation metric.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{F1-Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

$$\text{Sensitivity} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \tag{5}$$

$$\text{TPR} = \frac{TP}{TP + FN} \tag{6}$$

$$\text{FPR} = \frac{FP}{FP + TN} \tag{7}$$

$$\text{AUC} = \int_0^1 \text{TPR}\, d(\text{FPR}) \tag{8}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{9}$$

where y_i is the actual value and $\hat{y}_i$ is the predicted value. The less the RMSE, the better is the performance of the architecture.

The proposed network is implemented in TensorFlow 2.0 and Keras, using Python 3.11. The experiments are conducted on a system equipped with an NVIDIA RTX 3060 GPU, an Intel Core i7 (3.5 GHz) CPU, and 128 GB of RAM.

3.7 Explainable Artificial Intelligence

To aid in the diagnostic process, an explainable AI (XAI) system was developed to identify the regions of the MRI scan that the CNN is using to classify the image. It gets harder to comprehend the logic underlying AI models' judgments as they get more complicated and smart. This raises possible problems in the medical field, as AI systems have the ability to make judgments that might change people's lives. XAI is very important in the context of medical picture categorization (Adarsh et al., 2024). XAI can result in more accurate assessments about patient care and treatment plans by assisting medical practitioners in comprehending and interpreting the decisions made by AI systems. It is also crucial for ethical and regulatory grounds, as medical organizations and authorities frequently demand openness in the way AI systems make decisions in medical applications to guarantee adherence to rules and norms. Several explainable AI techniques, such as LIME, SHAP, and attention map were applied to improve the interpretability and transparency of DL models used for AD detection. In this experiment, we used Gradient Class Activation Mapping (GradCAM),

Score-CAM, Faster Score-CAM, and XGRADCAM. To get the output for the GradCAM, the feature importance with respect to each feature map C_f is obtained as:

$$W_f = \sum_x \sum_y \frac{\partial S_c}{\partial C_f} \tag{10}$$

The importance weight is calculated as:

$$H_f = \text{ReLU}\left(\sum_y \left(\frac{1}{N} W_f\right) \times C_f\right) \tag{11}$$

The score for class c is given by:

$$S_c = \sum_f \frac{1}{z} \sum_x \sum_y C_f \tag{12}$$

The output of the global average pooling is defined as:

$$G_f = \frac{1}{N} \sum_x \sum_y C_f \tag{13}$$

Rewriting the score for class c:

$$S_c = \sum_f W_f G_f \tag{14}$$

The gradients for the class score with respect to the feature maps are computed as:

$$\frac{\partial G_f}{\partial C_f} = \frac{1}{N} \tag{15}$$

Finally, summing both sides of the equation we get the output of GradCAM:

$$\frac{\partial S_c}{\partial G_f} = W_f \frac{\partial S_c}{\partial C_f} N \tag{16}$$

The simplified equation will be:

$$S_C = \text{ReLU}\left(\sum_{k=1}^{K} \alpha_k \cdot A_k\right) \tag{17}$$

where:

$$\alpha_k = \frac{1}{Z} \sum_{i,j} \frac{\partial y_c}{\partial A_k(i,j)} \tag{18}$$

The Score-CAM equation is written as:

$$S_c = \text{ReLU}\left(\sum_{k=1}^{K} \omega_k \cdot A_k\right) \tag{19}$$

Where the weight ω_k is calculated by:

$$\omega_k = \frac{\exp(f_k)}{\sum_{k'} \exp(f_{k'})} \tag{20}$$

The XGrad-CAM equation is written as:

$$S_c = \text{ReLU}\left(\sum_{k=1}^{K} \alpha_k^X \cdot A_k\right) \tag{21}$$

Where the weight α_k^X is given by:

$$\alpha_k^X = \frac{1}{Z} \sum_{i,j} \left(\frac{\partial y_c}{\partial A_k(i,j)} + \eta\right) \tag{22}$$

4 Experimentation and Analysis

4.1 Comparative Analysis

For our analysis, we use four datasets, each containing different sets of samples for various classes. To evaluate the performance of our model, we use several metrics: accuracy, precision, recall, F1-score, RMSE, AUC-score, and sensitivity. In our first experiment, we use the Kaggle dataset, which contains 6,400 brain image samples. Originally, it has four classes. We experiment on two multiclass classifications and one binary classification. For 4-class classification, we outperform existing approaches with an accuracy of 99.66%, an RMSE score of only 0.0547, and a sensitivity of 99.21%. For 3-class classification, we achieve an accuracy of 99.63%. Our proposed model achieves a perfect accuracy of 100% when predicting between two classes: demented or non-demented. The OASIS dataset, one of the largest datasets, is used for similar experiments with 4-class, 3-class, and 2-class classifications. We achieve very promising performance with accuracies of 99.92%, 99.90%, and 99.95%, respectively.

We also use two ADNI datasets in this study. For the ADNI-1 dataset, we conduct four experiments using our proposed architecture. We create datasets for three different brain regions—axial, coronal, and sagittal planes—and one combining all planes. Our architecture performs exceptionally well in all cases. For the axial plane, we achieve an accuracy of 99.08% with an AUC score of 99.80%. The sagittal plane yields the highest accuracy, with 99.85% correctly classifying Alzheimer's progression. For the coronal plane and the mixed-plane dataset, we achieve accuracies of 99.50% and 99.17%, respectively. For this dataset, 3-class classification is performed in all experiments.For the ADNI-2 dataset, we conduct two experiments on 5-class and 3-class classifications. For 5-class classification, our proposed method achieves an accuracy of 97.79%, and for 3-class classification, we achieve 98.60% accuracy.

When comparing our results with existing studies, we find that most studies using the ADNI and OASIS datasets use significantly smaller sample sizes. In contrast, our study achieves superior results and conducts a more extensive and detailed analysis. A comparison of results from existing literature is shown in Table 3, and the detailed results of our experiments are presented in Table 5.

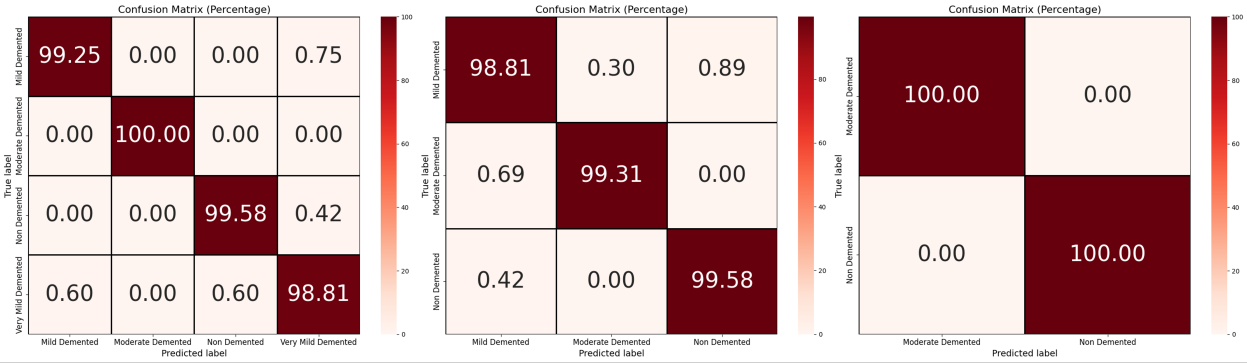
Table 3: Comparison of our Proposed Frameworks with other Recent Works

Paper	*Approach*	*Datasets*	*Data Type*	*Number of Classes*	*Number of Samples*	*Performance*			
						Accuracy	*Precision*	*Recall*	*F1-score*
Murugan et al. (2021)	CNN	Kaggle	MRI	4	6400 to 128000 using SMOTE Augment	95.23%	96%	95%	95.27%
Kaplan et al. (2021)	Local Phase Quantization Network	Kaggle	MRI	2 4	-	99.64% 99.62%	99.72% 99.74%	- 99.66%	99.64% 99.70%
Sharma et al. (2022)	Transfer Learning + Permutation Based Voting	Kaggle	MRI	4	6126	91.75%	-	-	90.25%
Al-Adhaileh (2022)	AlexNet	Kaggle	MRI	-	-	94.53%	-	-	94.12%
Ullah and Jamjoom (2023)	CNN	Kaggle	MRI	4	6400 to 128000 using SMOTE Augment	99.38%	99%	99%	99%
Biswas et al. (2021)	CNN	Kaggle	MRI	2	4800	99.38%	99.70%	95%	99.32%
De Santi et al. (2023)	3D CNN	ADNI	18F-FDG PET	2	2552	92%	-	-	-
Jahan et al. (2023b)	EfficientNet-B7	ADNI	MRI	5	10,025	96.34%	96%	96%	96%
Arafa et al. (2024)	CNN VGG16	Kaggle	MRI	2	6400	97.44%	97.46%	97.49%	97.48%
Noh et al. (2023)	3D-CNN- Phase LSTM	ADNI	fMRI	4	-	96.43%	-	-	-
Bamber and Vishvakarma (2023)	CNN	OASIS	MRI	4	-	98%	-	-	-
El-Assy et al. (2024)	CNN	ADNI	MRI	3 4	-	99.43% 99.57%	99.43% 99.57%	99.43% 99.57%	99.43% 99.57%
Dhaygude et al. (2024)	3D-CNN	ADNI	MRI	2	638	94.48%	-	-	-
Tripathy et al. (2024)	Depthwise Separarable CNN	OASIS Kaggle	MRI	4	399 6400	96.25% 99.75%	96.71% 99.63%	96.36% 99.77%	96.52% 99.99%
Mahim et al. (2024)	Vision Transformer + GRU	Kaggle ADNI	MRI	4 2 3	6400 6400 2970	99.53% 99.69% 99.26%	99.53% 99.69% 99.26%	99.53% 99.69% 99.26%	99.53% 99.69% 99.26%

Table 5: Results of our proposed method on four different datasets and with different numbers of classes

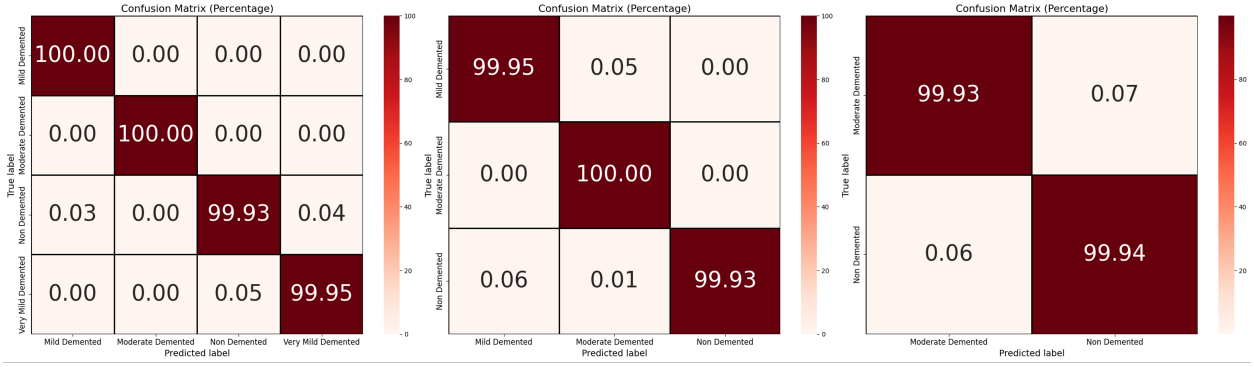
Dataset	*Classes*	*Total Samples*	*Results*						
			Accuracy	Precision	Recall	F1-score	RMSE	AUC	Sensitivity
Kaggle Dataset	4	6400	99.66%	99.66%	99.66%	99.66%	0.0547	97.12%	99.21%
	3		99.63%	99.63%	99.63%	99.63%	0.0157	99.81%	99.24%
	2		100%	100%	100%	100%	0.000607	99.80%	100%
OASIS	4	80,000	99.92%	99.92%	99.92%	99.92%	0.0157	99.14%	99.97%
	3		99.90%	99.90%	99.90%	99.90%	0.0241	99.88%	99.96%
	2		99.95%	99.95%	99.95%	99.95%	0.0202	99.92%	99.9.4%
ADNI-GO	5	18,775	97.79%	97.79%	97.79%	97.79%	0.0927	92.71%	89.94%
	3	18,151	98.60%	98.60%	98.60%	98.60%	0.0967	98.23%	95%
ADNI-1	3 (Axial Plane)	12,860	99.08%	99.08%	99.08%	99.08%	0.0634	99.80%	98.89%
	3 (Sagittal Plane)		99.85%	99.85%	99.85%	99.85%	0.0251	99.91%	99.58%
	3 (Coronal Plane)		99.5%	99.56%	99.56%	99.56%	0.0481	99.87%	99.35%
	3 (All Plane)	38,580	99.17%	99.17%	99.17%	99.17%	0.0805	99.48%	99.08%

We added visual aids to these numerical data to give a more comprehensive view of our investigations. In the figures below, we present the confusion matrix, offering a detailed breakdown of its classification performance. The confusion matrices in Figures 7, 8, 9, and 10 shows the detailed result overview of all experiments conducted with proposed approach with different classes for datasets: kaggle, oasis, adni-1 and adni-go respectively.



(a) (b) (c)

Figure 7: Confusion matrix of Kaggle Dataset: (a) 4 Classes (b) 3 classes (c) 2 classes



(a) (b) (c)

Figure 8: Confusion matrix of OASIS Dataset: (a) 4 Classes (b) 3 classes (c) 2 classes

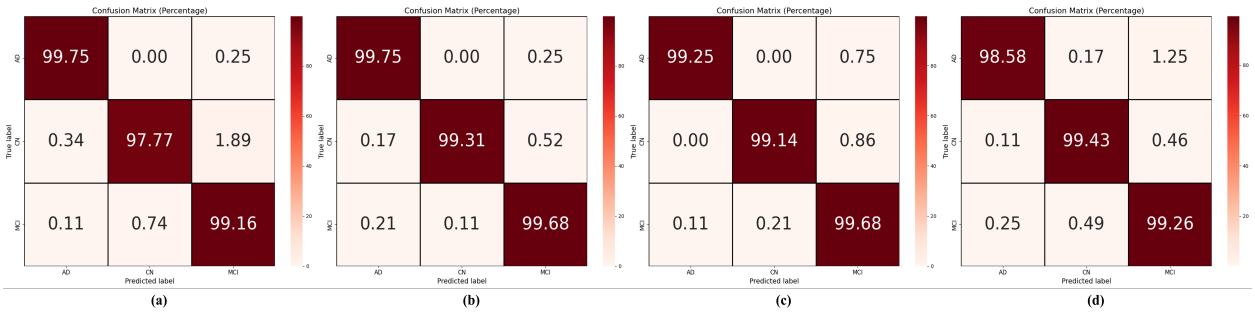


Figure 9: Confusion matrix of ADNI-1 Dataset: (a) 3 Classes-axial (b) 3 classes- sagittal (c) 3 classes- coronal (d) 3 classes- all planes

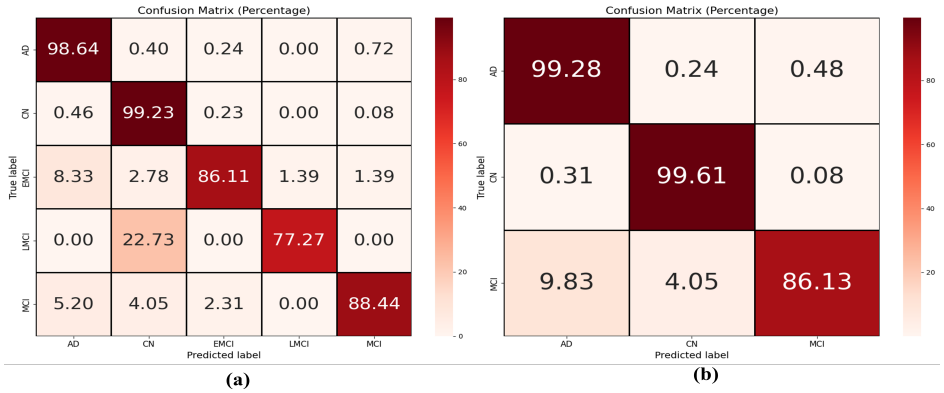


Figure 10: Confusion matrix of ADNI-GO Dataset: (a) 5 Classes (b) 3 classes

4.2 Explainability Analysis

The Figure 11 below, highlights areas of significance for decision-making by comparing saliency maps from several models applied to brain MRI scans. Concentrated red/yellow regions in the heatmaps show that the suggested model has superior focus, paying closer attention to areas that are diagnostically significant. Other models, on the other hand, tend to highlight unimportant regions and show more scattered or chaotic attention. This demonstrates how the proposed framework may more effectively capture clinically relevant aspects, improving its interpretability and usefulness in medical applications.

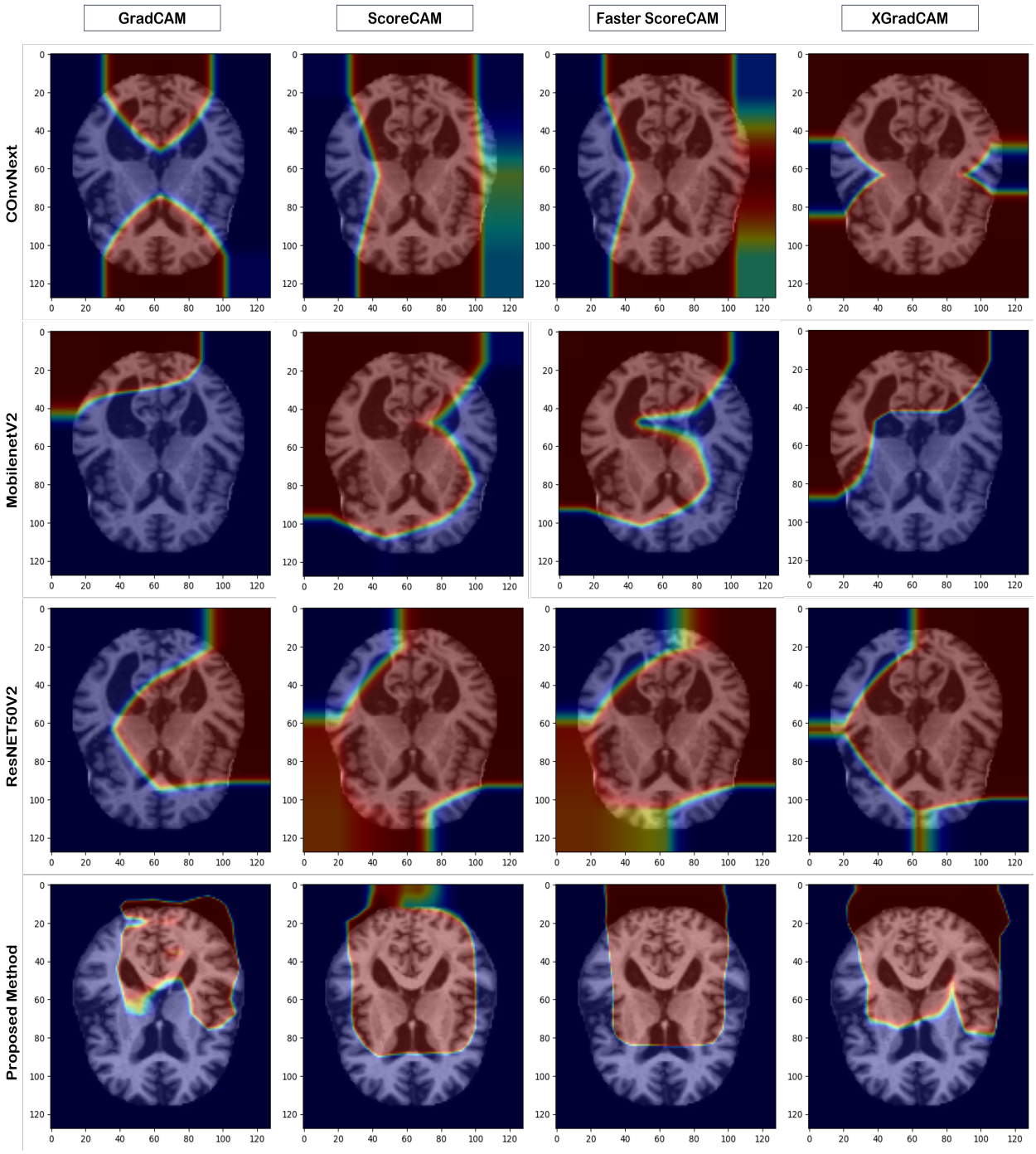


Figure 11: Explainability comparison of proposed method with other state-of-the-art methods

4.3 Ablation study

Understanding the contributions of various components to a machine learning model's performance requires comprehensive ablation studies. Within the framework of our CNN image categorization model, we conduct a detailed investigation to assess the impact of different hyperparameters and architectural choices. Prior to finalizing the hyperparameters, we perform extensive experiments by varying their values. For optimizer selection, we test Adam, SGD, and RMSProp optimizers with learning rates of 0.001, 0.005, 0.0001, 0.0005, and 0.00001. For learning rate schedulers, we evaluate ReduceLROnPlateau, Exponential Decay, and Cosine Annealing Decay. Based on these experiments, we conclude

that the Adam optimizer with an initial learning rate of 0.0001 is the optimal choice. We also explore different image preprocessing techniques and data augmentation ranges. For image processing, we experiment with original images, sharpened images, and various colormap applications. A sample image of preprocessing techniques is provided in Figure 12. Additionally, we test various network architectures by altering the number of layers and kernel sizes. To evaluate the model's performance under different conditions, we conduct experiments using both unbalanced datasets and balanced datasets with augmentation. These investigations provide insights into the impact of preprocessing, augmentation, and architectural design on model performance.

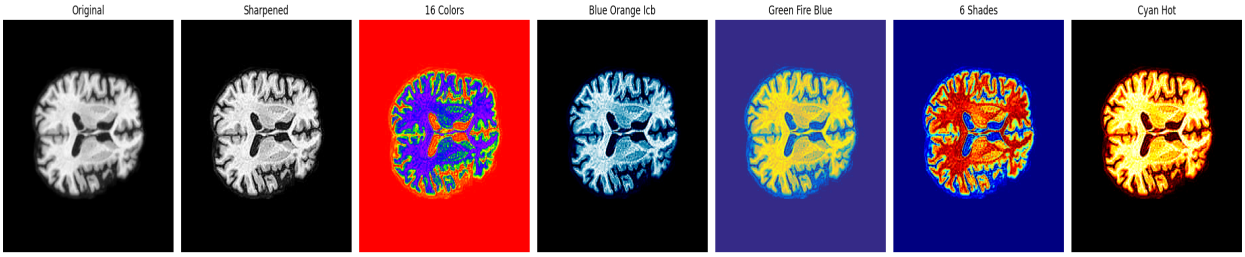


Figure 12: Different processed image: original, sharpened, Rainbow, Blue Orange Icb, Green Fire Blue, 6 Shades, Cyan Hot

5 Conclusion and Future Directions

Effective diagnosis of neurological disorders such as AD and MCI depends on prompt detection and treatments. By presenting a state-of-the-art solution that combines the transparency of XAI with the capability of sophisticated deep learning algorithms, the study described here helps to meet this pressing requirement. Our suggested architecture is both lightweight and sophisticated. The architecture includes two self-attention layers, the multihead-attention layer and group query attention, as well as a complicated spatial attention layer to increase emphasis on spatial and temporal elements. The proposed approach performs better across all four datasets and experimentation categories. The model's interpretability is further improved by the use of XAI approaches, which give doctors a better understanding of the decision-making process and the factors affecting AD diagnosis. Clinicians may improve treatment strategies and make well-informed judgments by visualizing the important features that contribute to the model's predictions. The suggested paradigm has the potential to diagnose AD early and accurately, allowing for swift responses and better patient care. Although our findings are encouraging, there is still much work to be done in order to diagnose cognitive diseases as accurately as possible. Subsequent versions of our study will explore the possible advantages of using different kernel functions to better combine multimodal data. In order to increase the range of our diagnosis, we are also eager to investigate the integration of several imaging modalities. The proposed architecture is highly suited for image-based approaches. For further enhancement, incorporating graph theory could offer valuable new insights into AD diagnosis. In conclusion, our research represents a significant step forward in advancing cognitive disease detection.

References

AbdulAzeem, Y., Bahgat, W. M., and Badawy, M. (2021). A cnn based framework for classification of alzheimer's disease. *Neural Computing and Applications*, 33(16):10415–10428.

Abuhmed, T., El-Sappagh, S., and Alonso, J. M. (2021). Robust hybrid deep learning models for alzheimer's progression detection. *Knowledge-Based Systems*, 213:106688.

Adarsh, V., Gangadharan, G., Fiore, U., and Zanetti, P. (2024). Multimodal classification of alzheimer's disease and mild cognitive impairment using custom mkscddl kernel over cnn with transparent decision-making for explainable diagnosis. *Scientific Reports*, 14(1):1774.

Ahanger, A. B., Aalam, S. W., Assad, A., Macha, M. A., and Bhat, M. R. (2024). Alzhinet: an explainable self-attention based classification model to detect alzheimer from 3d volumetric mri data. *International Journal of System Assurance Engineering and Management*, pages 1–10.

Al-Adhaileh, M. H. (2022). Diagnosis and classification of alzheimer's disease by using a convolution neural network algorithm. *Soft Computing*, 26(16):7751–7762.

Almohimeed, A., Saad, R. M., Mostafa, S., El-Rashidy, N., Farag, S., Gaballah, A., Abd Elaziz, M., El-Sappagh, S., and Saleh, H. (2023). Explainable artificial intelligence of multi-level stacking ensemble for detection of alzheimer's disease based on particle swarm optimization and the sub-scores of cognitive biomarkers. *IEEE Access*.

Ansingkar, N., Patil, R. B., and Deshmukh, P. (2022). An efficient multi class alzheimer detection using hybrid equilibrium optimizer with capsule auto encoder. *Multimedia Tools and Applications*, 81(5):6539–6570.

Arafa, D. A., Moustafa, H. E.-D., Ali, H. A., Ali-Eldin, A. M., and Saraya, S. F. (2024). A deep learning framework for early diagnosis of alzheimer's disease on mri images. *Multimedia Tools and Applications*, 83(2):3767–3799.

Association, A. (2019). 2019 alzheimer's disease facts and figures. *Alzheimer's & dementia*, 15(3):321–387.

Bamber, S. S. and Vishvakarma, T. (2023). Medical image classification for alzheimer's using a deep learning approach. *Journal of Engineering and Applied Science*, 70(1):54.

Bari Antor, M., Jamil, A. S., Mamtaz, M., Monirujjaman Khan, M., Aljahdali, S., Kaur, M., Singh, P., and Masud, M. (2021). A comparative analysis of machine learning algorithms to predict alzheimer's disease. *Journal of Healthcare Engineering*, 2021(1):9917919.

Beltran, J. F., Wahba, B. M., Hose, N., Shasha, D., Kline, R. P., and Initiative, A. D. N. (2020). Inexpensive, non-invasive biomarkers predict alzheimer transition using machine learning analysis of the alzheimer's disease neuroimaging (adni) database. *PloS one*, 15(7):e0235663.

Biswas, M., Mahbub, M. K., and Miah, M. A. M. (2021). An enhanced deep convolution neural network model to diagnose alzheimer's disease using brain magnetic resonance imaging. In *International Conference on Recent Trends in Image Processing and Pattern Recognition*, pages 42–52. Springer.

Böhle, M., Eitel, F., Weygandt, M., and Ritter, K. (2019). Layer-wise relevance propagation for explaining deep neural network decisions in mri-based alzheimer's disease classification. *Frontiers in aging neuroscience*, 11:456892.

Cole, J. H. and Franke, K. (2017). Predicting age using neuroimaging: innovative brain ageing biomarkers. *Trends in neurosciences*, 40(12):681–690.

De Santi, L. A., Pasini, E., Santarelli, M. F., Genovesi, D., and Positano, V. (2023). An explainable convolutional neural network for the early diagnosis of alzheimer's disease from 18f-fdg pet. *Journal of Digital Imaging*, 36(1):189–203.

Dhaygude, A. D., Ameta, G. K., Khan, I. R., Singh, P. P., Maaliw III, R. R., Lakshmaiya, N., Shabaz, M., Khan, M. A., Hussein, H. S., and Alshazly, H. (2024). Knowledge-based deep learning system for classifying alzheimer's disease for multi-task learning. *CAAI Transactions on Intelligence Technology*.

Ding, X., Bucholc, M., Wang, H., Glass, D. H., Wang, H., Clarke, D. H., Bjourson, A. J., Dowey, L. R. C., O'Kane, M., Prasad, G., et al. (2018). A hybrid computational approach for efficient alzheimer's disease classification based on heterogeneous data. *Scientific reports*, 8(1):9774.

Ebrahimi, A., Luo, S., and Disease Neuroimaging Initiative, f. t. A. (2021). Convolutional neural networks for alzheimer's disease detection on mri images. *Journal of Medical Imaging*, 8(2):024503–024503.

El-Assy, A., Amer, H. M., Ibrahim, H., and Mohamed, M. (2024). A novel cnn architecture for accurate early detection and classification of alzheimer's disease using mri data. *Scientific Reports*, 14(1):3463.

El-Sappagh, S., Alonso, J. M., Islam, S. R., Sultan, A. M., and Kwak, K. S. (2021). A multilayer multimodal detection and prediction model based on explainable artificial intelligence for alzheimer's disease. *Scientific reports*, 11(1):2660.

Ieracitano, C., Mammone, N., Hussain, A., and Morabito, F. C. (2020). A novel multi-modal machine learning based approach for automatic classification of eeg recordings in dementia. *Neural Networks*, 123:176–190.

Jahan, S., Abu Taher, K., Kaiser, M. S., Mahmud, M., Rahman, M. S., Hosen, A. S., and Ra, I.-H. (2023a). Explainable ai-based alzheimer's prediction and management using multimodal data. *Plos one*, 18(11):e0294253.

Jahan, S., Saif Adib, M. R., Mahmud, M., and Kaiser, M. S. (2023b). Comparison between explainable ai algorithms for alzheimer's disease prediction using efficientnet models. In *International conference on brain informatics*, pages 357–368. Springer.

Jain, R., Jain, N., Aggarwal, A., and Hemanth, D. J. (2019). Convolutional neural network based alzheimer's disease classification from magnetic resonance brain images. *Cognitive Systems Research*, 57:147–159.

Jin, D., Xu, J., Zhao, K., Hu, F., Yang, Z., Liu, B., Jiang, T., and Liu, Y. (2019). Attention-based 3d convolutional network for alzheimer's disease diagnosis and biomarkers exploration. In *2019 IEEE 16Th international symposium on biomedical imaging (ISBI 2019)*, pages 1047–1051. IEEE.

Kang, W., Li, B., Papma, J. M., Jiskoot, L. C., Deyn, P. P. D., Biessels, G. J., Claassen, J. A., Middelkoop, H. A., Flier, W. M. v. d., Ramakers, I. H., et al. (2023). An interpretable machine learning model with deep learning-based imaging biomarkers for diagnosis of alzheimer's disease. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 69–78. Springer.

Kaplan, E., Dogan, S., Tuncer, T., Baygin, M., and Altunisik, E. (2021). Feed-forward lpqnet based automatic alzheimer's disease detection model. *Computers in Biology and Medicine*, 137:104828.

Kavitha, C., Mani, V., Srividhya, S., Khalaf, O. I., and Tavera Romero, C. A. (2022). Early-stage alzheimer's disease prediction using machine learning models. *Frontiers in public health*, 10:853294.

Lahmiri, S. (2023). Integrating convolutional neural networks, knn, and bayesian optimization for efficient diagnosis of alzheimer's disease in magnetic resonance images. *Biomedical Signal Processing and Control*, 80:104375.

Lee, B., Ellahi, W., and Choi, J. Y. (2019a). Using deep cnn with data permutation scheme for classification of alzheimer's disease in structural magnetic resonance imaging (smri). *IEICE TRANSACTIONS on Information and Systems*, 102(7):1384–1395.

Lee, G., Nho, K., Kang, B., Sohn, K.-A., and Kim, D. (2019b). Predicting alzheimer's disease progression using multi-modal deep learning approach. *Scientific reports*, 9(1):1952.

Liu, Z., Lu, H., Pan, X., Xu, M., Lan, R., and Luo, X. (2022). Diagnosis of alzheimer's disease via an attention-based multi-scale convolutional neural network. *Knowledge-Based Systems*, 238:107942.

Mahim, S., Ali, M. S., Hasan, M. O., Nafi, A. A. N., Sadat, A., Al Hasan, S., Shareef, B., Ahsan, M. M., Islam, M. K., Miah, M. S., et al. (2024). Unlocking the potential of xai for improved alzheimer's disease detection and classification using a vit-gru model. *IEEE Access*.

Moser, E., Stadlbauer, A., Windischberger, C., Quick, H. H., and Ladd, M. E. (2009). Magnetic resonance imaging methodology. *European journal of nuclear medicine and molecular imaging*, 36:30–41.

Murugan, S., Venkatesan, C., Sumithra, M., Gao, X.-Z., Elakkiya, B., Akila, M., and Manoharan, S. (2021). Demnet: A deep learning model for early diagnosis of alzheimer diseases and dementia from mr images. *Ieee Access*, 9:90319–90329.

Neugroschl, J. and Wang, S. (2011). Alzheimer's disease: diagnosis and treatment across the spectrum of disease severity. *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine*, 78(4):596–612.

Noh, J.-H., Kim, J.-H., and Yang, H.-D. (2023). Classification of alzheimer's progression using fmri data. *Sensors*, 23(14):6330.

Park, H. Y., Shim, W. H., Suh, C. H., Heo, H., Oh, H. W., Kim, J., Sung, J., Lim, J.-S., Lee, J.-H., Kim, H. S., et al. (2023). Development and validation of an automatic classification algorithm for the diagnosis of alzheimer's disease using a high-performance interpretable deep learning network. *European Radiology*, 33(11):7992–8001.

Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., Jack Jr, C., Jagust, W. J., Shaw, L. M., Toga, A. W., et al. (2010). Alzheimer's disease neuroimaging initiative (adni) clinical characterization. *Neurology*, 74(3):201–209.

Pinamonti, M. (2021). Alzheimer MRI 4 classes dataset. `https://www.kaggle.com/datasets/marcopinamonti/alzheimer-mri-4-classes-dataset/data`. [Online; accessed 20-July-2024].

Rallabandi, V. S., Tulpule, K., Gattu, M., Initiative, A. D. N., et al. (2020). Automatic classification of cognitively normal, mild cognitive impairment and alzheimer's disease using structural mri analysis. *Informatics in Medicine Unlocked*, 18:100305.

Salehi, A. W., Baglat, P., Sharma, B. B., Gupta, G., and Upadhya, A. (2020). A cnn model: earlier diagnosis and classification of alzheimer disease using mri. In *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, pages 156–161. IEEE.

Scheltens, P., De Strooper, B., Kivipelto, M., Holstege, H., Chételat, G., Teunissen, C. E., Cummings, J., and van der Flier, W. M. (2021). Alzheimer's disease. *The Lancet*, 397(10284):1577–1590.

Shahbaz, M., Ali, S., Guergachi, A., Niazi, A., and Umer, A. (2019). Classification of alzheimer's disease using machine learning techniques. In *International Conference on Data Technologies and Applications*.

Shamrat, F. J. M., Akter, S., Azam, S., Karim, A., Ghosh, P., Tasnim, Z., Hasib, K. M., De Boer, F., and Ahmed, K. (2023). Alzheimernet: An effective deep learning based proposition for alzheimer's disease stages classification from functional brain changes in magnetic resonance images. *IEEE Access*, 11:16376–16395.

Sharma, S., Gupta, S., Gupta, D., Altameem, A., Saudagar, A. K. J., Poonia, R. C., and Nayak, S. R. (2022). Htlml: Hybrid ai based model for detection of alzheimer's disease. *Diagnostics*, 12(8):1833.

Tripathy, S. K., Nayak, R. K., Gadupa, K. S., Mishra, R. D., Patel, A. K., Satapathy, S. K., Bhoi, A. K., and Barsocchi, P. (2024). Alzheimer's disease detection via multiscale feature modelling using improved spatial attention guided depth separable cnn. *International Journal of Computational Intelligence Systems*, 17(1):113.

Ullah, Z. and Jamjoom, M. (2023). A deep learning for alzheimer's stages detection using brain images. *Computers, Materials & Continua*, 74(1).

Wang, C., Tachimori, H., Yamaguchi, H., Sekiguchi, A., Li, Y., Yamashita, Y., and Initiative, A. D. N. (2024). A multimodal deep learning approach for the prediction of cognitive decline and its effectiveness in clinical trials for alzheimer's disease. *Translational psychiatry*, 14(1):105.

Wu, Y., Zhou, Y., Zeng, W., Qian, Q., and Song, M. (2022). An attention-based 3d cnn with multi-scale integration block for alzheimer's disease classification. *IEEE Journal of Biomedical and Health Informatics*, 26(11):5665–5673.

Zhang, D., Shen, D., Initiative, A. D. N., et al. (2012). Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in alzheimer's disease. *NeuroImage*, 59(2):895–907.

Zhang, J., Zheng, B., Gao, A., Feng, X., Liang, D., and Long, X. (2021a). A 3d densely connected convolution neural network with connection-wise attention mechanism for alzheimer's disease classification. *Magnetic Resonance Imaging*, 78:119–126.

Zhang, X., Han, L., Zhu, W., Sun, L., and Zhang, D. (2021b). An explainable 3d residual self-attention deep neural network for joint atrophy localization and alzheimer's disease diagnosis using structural mri. *IEEE journal of biomedical and health informatics*, 26(11):5289–5297.

Zhang, Y. and Yang, Q. (2021). A survey on multi-task learning. *IEEE transactions on knowledge and data engineering*, 34(12):5586–5609.