

RESEARCH ARTICLE

AlzheimerNet: An Effective Deep Learning Based Proposition for Alzheimer's Disease Stages Classification From Functional Brain Changes in Magnetic Resonance Images

F M JAVED MEHEDI SHAMRAT¹, SHAMIMA AKTER², SAMI AZAM³, ASIF KARIM³, (Member, IEEE), PRONAB GHOSH⁴, ZARRIN TASNIM¹, KHAN MD. HASIB⁵, (Member, IEEE), FRISO DE BOER³, AND KAWSAR AHMED^{6,7}, (Graduate Student Member, IEEE)

¹Department of Software Engineering, Daffodil International University, Dhaka 1207, Bangladesh

²Department of Bioinformatics and Computational Biology, George Mason University, Fairfax, VA 22030, USA

³Faculty of Science and Technology, Charles Darwin University, Casuarina, NT 0909, Australia

⁴Department of Computer Science, Lakehead University, Thunder Bay, ON P7B 5E1, Canada

⁵Department of Computer Science and Engineering, Ahsanullah University of Science and Technology, Dhaka 1208, Bangladesh

⁶Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada

⁷Group of BiophotomatiX, Department of Information and Communication Technology, Mawlana Bhashani Science and Technology University, Santosh, Tangail 1902, Bangladesh

Corresponding author: Asif Karim (asif.karim@cdu.edu.au)

ABSTRACT Alzheimer's disease is largely the underlying cause of dementia due to its progressive neurodegenerative nature among the elderly. The disease can be divided into five stages: Subjective Memory Concern (SMC), Mild Cognitive Impairment (MCI), Early MCI (EMCI), Late MCI (LMCI), and Alzheimer's Disease (AD). Alzheimer's disease is conventionally diagnosed using an MRI scan of the brain. In this research, we propose a fine-tuned convolutional neural network (CNN) classifier called AlzheimerNet, which can identify all five stages of Alzheimer's disease and the Normal Control (NC) class. The ADNI database's MRI scan dataset is obtained for use in training and testing the proposed model. To prepare the raw data for analysis, we applied the CLAHE image enhancement method. Data augmentation was used to remedy the unbalanced nature of the dataset and the resultant dataset consisted of 60000 image data on the 6 classes. Initially, five existing models including VGG16, MobileNetV2, AlexNet, ResNet50 and InceptionV3 were trained and tested to achieve test accuracies of 78.84%, 86.85%, 78.87%, 80.98% and 96.31% respectively. Since InceptionV3 provides the highest accuracy, this model is later modified to design the AlzheimerNet using RMSprop optimizer and learning rate 0.00001 to achieve the highest test accuracy of 98.67%. The five pre-trained models and the proposed fine-tuned model were compared in terms of various performance matrices to demonstrate whether the AlzheimerNet model is in fact performing better in classifying and detecting the six classes. An ablation study shows the hyperparameters used in the experiment. The suggested model outperforms the traditional methods for classifying Alzheimer's disease stages from brain MRI, as measured by a two-tailed Wilcoxon signed-rank test, with a significance of < 0.05.

INDEX TERMS Alzheimer's disease, AlzheimerNet, deep learning, MRI imaging, multiclassification.

The associate editor coordinating the review of this manuscript and approving it for publication was Mehul S. Raval¹.

I. INTRODUCTION

Alzheimer's disease is a progressive neurodegenerative health condition that causes dementia, damage to brain cells

and cognitive impairment over time among patients [1], [2]. Detecting Alzheimer's is a difficult and time-consuming process, taking an average of 2.8 years in cases of late-onset and 4.4 years in cases of early-onset dementia [3]. A few years are just a little period of opportunity to improve the diagnostic method. There is considerable evidence that AD is the most prevalent kind of dementia that begins to show clinical indications at the age of 65 or above [4]. It is vital to advance diagnosis in order to speed up the discovery of a cure because late therapy is likely to be a key factor in failed treatment [5]. Furthermore, primary and correct diagnosis has the considerable possibility to minimize health care expenditures since they provide patients with supportive care that enables them to avoid incarceration [6]. There have been and continue to be technological advancements in the field of medical imaging that assess diagnostic medical photographs and provide in-depth information about the diagnostic to clinicians, researchers, and academics to inspire more study and analysis [7], [8]. Medical imaging techniques used today include fMRI, MRI, CT, EEG, and PET scans, among others. Many studies have been conducted using MRI brain scans. The MRI-scanned data have a substantial benefit in that they provide better spatial resolutions and that image features are visible to diagnose disease. AD most usually affects the elderly. Its prevalence is predicted to treble within the coming two decades as that of the ageing population, posing a substantial burden on society [9].

One of the most obvious characteristics of AD disease is brain shrinkage spreading from AD hallmark regions throughout the whole cortical region, as seen by an MRI scan. These apparent structural alterations occurred long before there was a significant loss in cognition, providing a potential for an early diagnosis of AD utilizing imaging techniques. To stop the aberrant deterioration of brain tissue, save health-care costs, and enhance treatment, researchers are examining methods for detecting this disease sooner rather than later. Recent struggles in AD studies may emphasize the need for early detection and treatments [10]. Dementia diagnosis has come to rely heavily on neuroimaging techniques, and as a result, there have been a variety of newer diagnoses developed. Using machine learning, neuroimaging improves diagnostic accuracy for several kinds of dementia. Early identification and automated categorization of AD have recently evolved, resulting in large-scale multimodal neuroimaging data. MRI, PET, and genetic sequencing findings are among the several techniques used to examine AD.

The ADNI database is one of the most credible sources of brain imaging data [11]. The database contains weighted MRI scans of patients at different stages in the progress of AD. The dataset contains six classes: AD, CN, MCI, EMCI, LMCI, and SMC. Mild cognitive impairment (MCI) is the stage that occurs between the normal ageing-related decrease in memory and thinking and the more severe dementia-related decline. The trend is slower in EMCI patients, who reflect those with moderate levels of cognitive and functional impairment. The rate of development of dementia and issues with

language or judgment was greater in LMCI groups. Utilizing deep learning models, the features of the disease can easily be identified for faster and early diagnosis. The proposed model AlzheimerNet structure employs numerous layers to assess functional and structural MRI data for the recognition of Alzheimer's disease. This differs from other comparable models, which often employ only a single type of network or analyze only structural or functional MRI data. Additionally, AlzheimerNet can perform on a high volume of data. It implements a multi-task learning strategy in which the network concurrently learns to identify both the progression and diagnosis of Alzheimer's disease, whereas previous models normally only predict one or the other. The key contribution of the research work are as follows:

- Exploring filters for smoothing the images.
- Exploring high volume dataset; this is new for Alzheimer's disease classification.
- The proposed method integrates extracted features with deep learning in a novel way for the classification of Alzheimer's disease from the ADNI dataset.

Following is how this paper was processed:

- A fine-tuned AlzheimerNet model is proposed in this work.
- The model is able to classify individuals into the five phases of Alzheimer's disease namely: MCI, EMCI, LMCI, SMC, AD, and overall six classes, including NC.
- The performance of five different CNN models, VGG16, ResNet50, MobileNetV2, AlexNet, and InceptionV3, is compared to AlzheimerNet to determine the proposed model's efficiency in performance..
- Data pre-processing along with augmentation are carried out to enhance the features and balance the dataset.

The following steps are taken for performance comparison and model evaluation:

- Parameters like accuracy, recall, precision, f1-score, etc. are computed from the confusion matrix and used to evaluate the models' efficiency.
- Grad-CAM is designed to provide a visual representation of the model.
- To evaluate the proposed model, a two-tailed Wilcoxon signed-rank p-value is worked out.
- To further evaluate the proposed model, the ROC curve is generated for the model and the AUC value is calculated.

The remaining contents of this article will be presented in the following format. Section II contains the literature review of research done previously for AD detection using the ADNI database with the employed models' accuracy rates. Section III describes the dataset collection, pre-processing and preparation before being fed into the model. Section IV illustrated the selected transfer learning models, their parameters and the proposed model. Section V portrayed the ablation study to choose the hyperparameters for the fine-tuned model (AlzheimerNet). Section VI discusses the result and the discussion. Section VII summarizes the study and discusses

further research. Figure 1 depicts the entire research process flow.

II. LITERATURE REVIEW

Machine learning provides a method for automatic categorization by learning complicated and nuanced patterns derived from massive volumes of data. In AD studies, various techniques are developed on a regular basis to automate diagnosis and forecast prospective medical state utilizing biomarkers.

To classify AD using MRI and PET brain scans, research [12], suggests the use of a cascaded CNN model. Multiple imaging modalities may be used to automatically learn the general, multi-level, and multi-modal characteristics for categorization, which are rather resistant to changes in scale and rotation. The pre-processing of the brain scans does not need any image segmentation or rigorous registration. A total of 397 people, including 93 Alzheimer's patients, 204 people with mild cognitive impairment, and 100 normal controls (NC), were scanned as part of the ADNI database. Based on the experimentation results, the suggested strategy has a 93.26% accuracy rate. This study [13] tries to demonstrate a classifying approach that depends on RF (Random Forest) feature selection and CNN classification. The model is trained on a mixed cohort of patients with HC, AD, MCI, and cMCI. Overall, this effort had a 38.8% accuracy rate.

The authors in [14] propose three effective methods for creating graphical descriptions for AD classification using 3DCNN. The 3D-VGGNet and 3D-ResNet methods for categorizing MRI scans from the AD and normal cohorts (NC) are trained using ADNI brain MRI images in this study. The area under the receiver operating characteristic curve (AUC) and classification accuracy (ACC) are used to quantify classification performance, with 3D-ResNet achieving the best accuracy of 79.4%. The authors of [15] suggest a technique for diagnosing AD that uses a CNN and a mix of sMRI and DTI techniques on a hippocampus ROI using data from the ADNI database. The authors selected 214 individuals, 48 AD cases, 108 MCI cases, as well as 58 NC patients. Each patient is scanned with a T1-weighted sMRI and a DTI scan. Additionally, the authors provide an approach for balancing classes of varied sizes and investigate the influence of ROI size on classification results, achieving a classification accuracy of 96.7%. This paper's [16] goal is to evaluate the accuracy of MCI to AD conversion using MRI data and a deep learning method based on CNNs. To begin, MRI images are processed for age correction. Second, local patches are taken from these scans and combined into 2.5 dimensions. The segments from AD and NC are then used to train a CNN to identify the deep learning features of MCI patients. Following that, FreeSurfer is used to mine structural brain picture characteristics to aid CNN. Finally, all sets of attributes are fed into a Learning algorithms classifier, which is used to forecast AD conversion. This method yields a 79.9% accuracy.

The purpose of this research [17] was to detect MCI and AD in 3089 T1-weighted MRI data from the ADNI-2 cohort and to examine its viability for predicting AD in the ADNI-1 cohort and an Asian cohort using a spectral graph-CNN with cortical thickness and shape. The model outperformed other deep learning approaches such as Euclidean domain-based multilayer networks and 1D CNNs on brain structure, as well as 2D and 3D CNNs on T1-weighted MRI data from the ADNI-2 cohort. On the ADNI-1 cohort, the fine-tuned graph-CNN achieved an excellent 89.4% accuracy of classification for CN vs AD. The authors developed a three-dimensional CNN for the purpose of detecting AD in [18]. The suggested model is based on 1230 PET scans of 988 individuals (169 AD, 661 MCI, and 400 CN). The ADNI database was used to acquire the data. The raw scans were stripped and normalized to exclude non-cerebral structures such as the skull, scalp, and dura, which reduces computational complexity and time. For CN/AD classification, the network obtains a similar accuracy of 88.76%. The authors of [19] extract feature using ensembles of basic CNN's then classify them using SoftMax cross-entropy. Then, they used a patch-based method in light of the data shortage. After preprocessing, authors concentrated on both the left and right hippocampal sections and input three view patches (TVPs) to the CNN, achieving 90.05% accuracy. The authors of [20] describe a deep learning technique for predicting the progression of AD in a specific patient using resting state fMRI data and a 3D CNN. This approach extracts spatial information from a four-dimensional volume and eliminates the time-consuming processes connected with extracting features in the past. In the experiment, a very basic deep learning architecture achieves 94.58% accuracy in classifying AD. This highlights the value of geometric feature extraction and classifiers for the identification of biomarkers for neurological diseases. The research [21] aimed to evaluate a novel deep-learning strategy for diagnosing AD and simultaneously predict Mental Status Assessment outcomes. The authors created operational 3D independent element spatial maps to implement as features in regression and classification problems by analyzing 331 individuals' resting-state fMRI images. For the classification job, a 3D CNN architecture was designed. In order to predict MMSE scores, we used linear least square, support vector, and bagging-based ensemble regressions, all of which were based on variables from the groups' independent component analyses. Techniques for maximizing features, such as the least absolute shrinkage and selection operator, as well as SVM-RFE, were employed to improve the performance of the MMSE regression. For the categorization of AD vs healthy controls, the mean balanced test accuracy was determined at 85.27%. By dealing directly with four-dimensional fMRI data, the authors of [22] describe a four-dimensional deep learning algorithm (C3d-LSTM) for AD classification that can quickly use spatial information. To extract spatial information from every region in a three-dimensional static picture sequence collected from fMRI, the proposed C3d-LSTM model integrates a number of 3D CNN models.

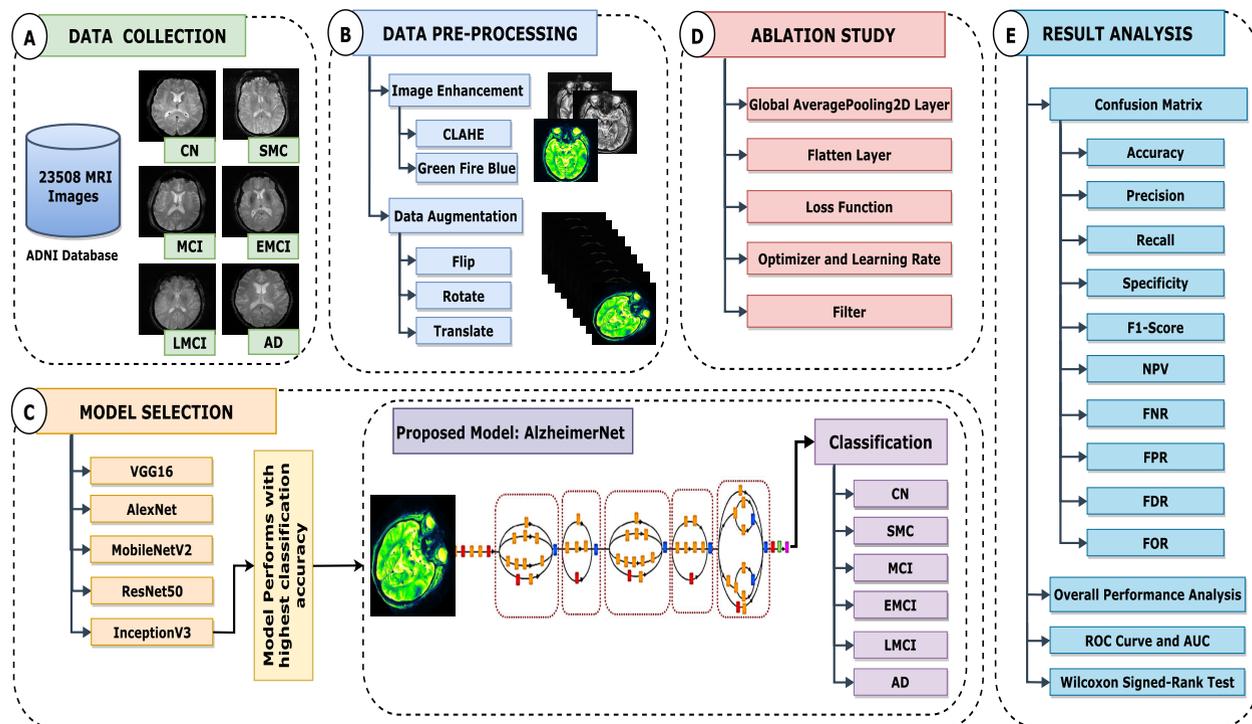


FIGURE 1. Working procedure of the entire proposed system (AlzheimerNet) for Multi-classification of Alzheimer's disease.

The instant information contained in the data was then retrieved using the produced features and supplied into an LSTM technique. The findings demonstrate the suitability of the suggested C3d-LSTM model for processing four-dimensional fMRI data and precisely extracting its spatiotemporal aspects for AD diagnosis. The primary goal of [23] is to apply a streamlined approach for feature extraction and classification of AD to fMRI data using a modified 3D CNN. The fMRI data were motion-corrected, normalized, and coregistered, the global mean and temporal signal drift was decreased, and intensity-based thresholding and masking were conducted before being input to the CNN. The trained network has a 93% success rate in properly labeling regions of 4D BOLD fMRI data as having either AD, cerebrovascular disease, encephalomyelitis, or traumatic brain injury.

In [24] machine learning techniques such as KNN, SVM, DT, LDA, and RF are used to classify the data (RF). Additionally, a novel CNN architecture is proposed for classifying the severity of AD. The objective of this research is to examine the link between Alzheimer's patients' functional MRI scans and their MMSE scores. The robust multitask feature learning technique is used to extract the features. Additionally, the Mini-Mental State Examination score is used to determine the severity, which is categorized as low, mild, moderate, or severe. The given CNN approach yields a 96.7% accuracy. The study in [25] proposes and evaluates a number of deep models and architectures, including two- and three-dimensional CNN and RNN. To deploy a 2D CNN to 3D MRI volumes, every MRI scan is split

into two 2D slices, with no respect for the interconnections among the 2 two-dimensional slices inside the MRI volume. Alternatively, a CNN model could be preceded by a Recurrent neural network in such a way that the two-dimensional CNN + RNN model can understand the connection across sequences of two-dimensional slices acquired during an MRI. The 3D voxel-based technique combined with transfer learning achieved a classification accuracy of 96.88%. The suggested approach in [26] creates multilayer perceptron representations of individual AD risk and high-resolution disease probability maps from regional brain activity. To prevent inequalities between classes, samples should be dispersed uniformly. The five subtypes of the ADNI dataset are called AD, MCI, EMCI, LMCI, and NC. The 1296 photos in the ADNI dataset were modified to fit the DEMNET model at 176×176 . The accuracy of the DEMNET model is 84.83%. The authors of [27] trained a CNN to identify AD utilizing $N = 663$ T1-weighted MRI scans of persons with dementia and amnesic MCI. The models were subsequently verified utilizing cross-validation and three individual datasets totalling $N = 1655$ cases. They examined the relationship between relevance ratings and hippocampal volume in order to demonstrate the approach's therapeutic value. To help with model understanding, the authors created an interactive depiction of 3D CNN significance maps that enables intuitive model evaluation. Classification of AD vs. CN was performed with a 94.9% accuracy.

According to the proposed model [28], which uses AlexNet and MRI datasets, a maximum of 95% and 0.1643% accuracy

and loss may be achieved for distinguishing non-demented, very mild, moderately severe dementing states. The AlexNet architecture combined with the Adam optimizer is being used to train MRI images using a variety of learning rates, including 0.0001, 0.001, 0.01, and 0.1, as well as loss binary cross-entropy. In [29], Angkoso et al. suggested a Multiplane CNN method on 1500 MRI Data from the ADNI dataset to classify AD, MCI and NC. The method uses brain extract tool (BET2) to remove the non-brain area from the MRI image. The proposed architecture is based on sequential CNN to identify the spatial structure data. The experiment achieved 93% overall accuracy in the classification of the three classes. A 2DCNN is proposed to classify AD and MCI using 3312 MRI images in [30]. The study uses BET2 for skull stripping in pre-processing the images. The proposed model is based on LNet-5 with the modification of the Leaky ReLU activation function and sigmoid function. Additionally, batch normalization was used to stabilize the learning process. The highest accuracy of 84% was achieved by the fine-tuned model successfully in the classification of AD. In [31], authors proposed a finetuned ResNet18 model to classify MCI AD and CN from MRI and PET data. The fine-tuned model used transfer learning and weighted loss function to balance the weight for each class. Additionally, the mish activation function was employed to increase classification accuracy. Overall, the proposed model can successfully classify with an accuracy of 88.3%. The authors propose a 3D CNN with dilated convolutions at the subject level in [32]. The system is designed using ADNI data from individuals with stable AD-kind dementia (sDAT) and normal control (sNC). The sDAT and sNC were identified from one other with an accuracy of 88% after a 5-fold cross-validation.

In this study in [33], authors used MRI to solve the job of AD multiclass classification while experimenting with randomized weights of ResNet18 and DenseNet201. The image's discriminating region was identified using a gradient class activation map for the suggested model's prediction. The experimental research revealed that the modified model could perform multiclass classification with an accuracy score of 98.89% suggesting that Alzheimer's disease may be classified using cutting-edge deep learning methods. The authors combined Resnet18 and Densenet121 to create a ResD hybrid technique that was utilized to classify Alzheimer's disease into five categories in [34]. For the study uses MRI data from the ADNI datastore. In this study [35], authors developed and tested CNN models for feature extraction of MRI image data and classification of AD. The model is then assessed using the validated model by examining the features that CNN (ResNet) retrieved from the fully connected layer. SVM, RF, and Softmax are the three ML classifiers that are used for each set of features and to analyze the outcomes. The ADNI and MIRIAD public datasets are used in this work. The classification accuracy produced by the combined models ResNet50 + SVM and ResNet50 + RF is 92% and 85.7%, respectively.

This study [36] will make use of the pre-trained models SqueezeNet, ResNet18, AlexNet, VGG16, DenseNet, and InceptionV3 to classify AD into four categories. The OASIS database's MRI images were used in the authors' experiments. After evaluating the implemented models, it is observed that the pre-trained SqueezeNet model derived the highest validation accuracy of 82.53% or multiclass classification of AD. For AD classification, authors in [37] suggested an improved VGG-16 architecture employing the arithmetic optimization algorithm (AOA). During pre-processing, the T1-weighted MRI image file is processed using the CAT12 toolbox. While the linear contrast stretching raises the picture contrast level. Similarly, image-enhancing techniques correct the unequal light distribution. After data preprocessing, an Optimized VGG-16 employing AOA successfully categorizes the four AD classes with 97.89% accuracy. The study [38] is an overview of various pre-trained CNN models applied of the ADNI dataset. The authors use MRI brain image to determine NC, MCI and AD classes. To determine the 3 classes, CNN models such as GoogLeNet, ResNet50, AlexNet, ResNet18, DenseNet and SqueezeNet were implemented. It is determined that the highest classification accuracy is obtained from GoogLeNet with 96.81% followed by ALexNet and ResNet50 with accuracies of 96.59% and 96.13% respectively. The neural networks used in the article [39] were applied to datasets of MRI images. There are five different types of subjects with Alzheimer's disease used in the study. For the classification challenge, authors employed MobileNet for transfer learning. For multi-class AD stage classifications, the MobileNet pre-trained model has been improved and has obtained 96.6% accuracy. By adopting a Learnable Weighted Pooling (LWP) method to transform 3D brain imaging into 2D fused images, the authors in [40] developed a unique 3D-2D methodology. The suggested model can quickly transfer the fused 2D picture by a pre-trained 2D model using the 3D-to-2D conversion while attaining higher performance over various 3D and 2D baselines. ResNet34 was used for feature extraction since it surpassed other 2D CNN structures and gains an accuracy of 88%.

III. DATASET

A. ADNI DATASET

The research utilized the Alzheimer's Disease Neuroimaging Initiative's (ADNI) dataset [11]. Michael Weiner, MD in 2003 founded ADNI as a public-private collaboration with the sole focus of testing whether MRI and PET scans, biological markers, biochemical tests, and evaluations of neurophysiology may be used to monitor the development of AD at every stage. The database provided 23508 T2-weighted MRI scans from 2456 different subjects. T2w scans were obtained in 3D using an FSE sequence with 3.0 mm slice thickness, 56 axial slices spanning the whole brain, a 256×256 scan matrix, a voxel size of 0.8594 × 0.8594 × 5 mm, with TR = 3000 ms, and TE = 95.2 ms for T2w. The imaging scans are divided into six categories namely NC, MCI, EMCI,

LMCI, SMC, and AD. A sample of the dataset is presented in Figure 2. Each scan consists of multiple slices that show the brain’s functional properties. Table 1 shows the demographics of the dataset utilized in the study. In this study, we included all subjects and their scan images however we excluded the variables “Sex” and “Age” as they do not contribute to the image processing classification task.

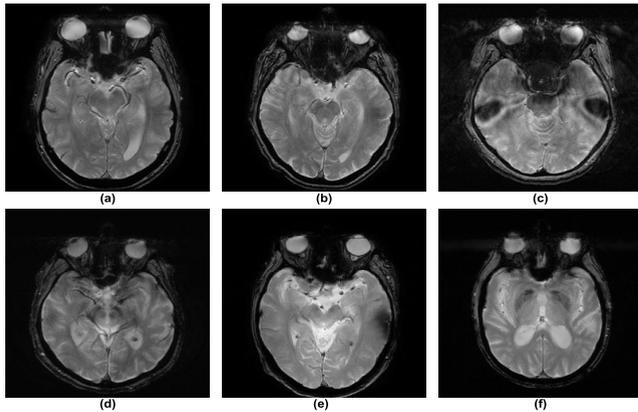


FIGURE 2. ADNI MRI dataset contains six classes, (a) NC, (b) MCI, (c) EMCI, (d) LMCI, (e) SMC, and (f) AD.

TABLE 1. Demographic information of the dataset.

Parameters	NC	MCI	EMCI	LMCI	SMC	AD
Subjects	801	617	333	178	111	416
Image	8125	5661	3946	2102	1100	2574
Sex (M/F)	3680/4445	3599/2062	2312/1634	1145/957	428/672	1441/1133
Age (years)	51-96	55-97	55-93	55-94	60-92	55-93
Mean (Age)	75.49	75.41	73.43	73.68	74.85	75.75
Standard Deviation (Age)	7.36	7.76	7.36	7.38	5.67	7.97

B. DATASET PREPARATION

The ADNI MRI data is collected in (Digital Imaging and Communication in Medicine) format. Since the CNN model employed in the detection requires image data in JPEG format, the 3D DICOM image is converted to 2D JPEG image. Using the DICOM converter windows application, the downloaded data from the ADNI website is converted to JPEG image. These converted data are viewable with any image viewer application are ready to be used as input in CNN models. In order to improve image quality, image enhancement and data augmentation approaches are used. This section describes the image processing techniques in detail.

1) IMAGE ENHANCEMENT

Image enhancement is a process that emphasizes certain parts of an image while simultaneously reducing or eliminating any unnecessary elements. For instance, reducing noise, revealing blurred features, and adjusting the intensity of a picture to attract attention to specific regions. Images are enhanced so that they may be processed or analyzed further by

increasing their contrasts and sharpness. Prior to processing, picture enhancement is the technique of boosting the original data’s quality and information content. Before being further processed, image enhancement increases the brightness and information of raw data. The enhancement doesn’t really increase the data’s fundamental information value; rather, it increases the contrast ratio of the selected look, making it more clearly recognizable [41].

a: APPLYING CONTRAST LIMITED ADAPTIVE HISTOGRAM EQUALIZATION (CLAHE)

CLAHE is a more sophisticated form of Adaptive Histogram Equalization (AHE) that outperforms AHE in improving the quality of complicated structures, as demonstrated by the authors in [42]. Additionally, it boosts local contrast to improve the utility of medical imaging [43] and works on tiny areas of an image to eliminate the false borders by combining the adjacent tiles utilizing bilinear interpolation [44], [45].

Suppose an image has a size of $N \times N$ and each tile in that image has a size of $n \times n$. Then, using Equation (1) the total number of tiles is computed,

$$T = \frac{N \times N}{n \times n} \tag{1}$$

Required histograms of the mentioned tiles are generated by leveraging the clip limit, C_L as mentioned in equation (2).

$$C_L = N_{CL} \times N_{AVG} \tag{2}$$

Here, N_{CL} refers to the Normalized form of Contrast Limit. N_{AVG} indicates the Number of average counts of the pixels in the image. Equation (3) is enforced to determine the value of N_{AVG} .

$$N_{AVG} = \frac{(N_x \times N_y)}{N_g} \tag{3}$$

From the equation, N_g , N_x and N_y refers to the Number of Gray levels, the number of pixels in x and y dimension respectively in the tile. Equation (4) calculates clipped pixel average counts.

$$N_{cp} = \frac{N \sum C_L}{N_g} \tag{4}$$

Here, C_L indicates clipped pixels and $N \sum C_L$ refers to the total number of C_L . The N_{cp} is dispersed among each corresponding gray level. However, some pixels remain to be distributed. To redistribute the remaining pixels, equation (5) is used.

$$R = \frac{N_g}{N_r} \tag{5}$$

where redistribution, R is computed by dividing N_g by the Number of pixels to be redistributed, N_r . Bi-linear interpolation merges neighborhood tiles to remove image data boundaries.

The CLAHE technique overcomes the constraints of global methods by enhancing local contrast. This technique, however, is dependent on two critical hyperparameters: (1) the

number of tiles and (2) the clip limit. An incorrect hyperparameter selection may significantly reduce image quality. Iteratively testing different values for these parameters leads to the optimal solution, which in this case is tileGridSize (12, 12), clip limit (3). Figure 3 illustrates the findings.

b: FILTER SELECTION

Applying a filter is the last enhancement process of this study. To “filter” an image means to alter its visual appearance by adjusting the pixel colors in the image. Applying filters has the effect of increasing contrast and adding a number of distinctive effects to the images. The properties of the image are increased and made more comparable to the model that is being trained, validated, and tested using this filter approach. It can be observed that from figure 4, several filters (Blue Orange Icb, 16 Colors, Green Fire Blue, 6 Shades, Cyan Hot) are applied to MRI scans images and selected the best filter with the highest accuracy. In Figure 5, the enhanced image is shown as output following the use of CLAHE and Green Fire Blue methods.

TABLE 2. Parameters used for data augmentation.

Augmentation Techniques	Ranges
Horizontal_flip	True
Vertical_flip	True
Rotate	45°
Rotate_horizontal	45°
Rotate_left	90°
Rotate_right	90°
Translate	x, y (18.2, 4.5)

C. IMAGE AUGMENTATION

A deep learning model needs huge data to perform properly. Some augmentation approaches are applied in this research to augment the enhanced data. By adding more diverse training dataset samples, data augmentation can help machine learning techniques perform better and produce better results. The algorithm performs well and is more reliable if the dataset utilized to train it is sufficiently vast and diversified. The use of data augmentation methods [46] increases the accuracy of the results. Additionally, data augmentation methods are a very well strategy for increasing the variety of the dataset. This strategy reduces overfitting in CNN models [47], [48] by providing sufficient diverse data for training. Flipping, rotating, zooming, mirroring, and cropping techniques are the most often used to augment data. In this study, seven augmentation techniques are utilized including horizontal flip, vertical flip, rotate 45°, rotate 45°horizontal, rotate 90° left, rotate 90° right, and translate on pre-processed data. Table 2 provides a concise overview of all available augmented settings. The dataset was expanded to a larger dataset containing 60000 images (NC: 10000, MCI: 10000, EMCI: 10000, LMCI: 10000, SMC: 10000, AD: 10000) after augmentation. The process of data augmentation is shown in figure 6.

The final step before training is to divide the dataset. The MRI scans were separated into three groups based on a 60:20:20 proportion for the training, validation, and test sets, respectively. After splitting the larger MRI Scans dataset of 60000 images into three subsets, the training set contains 36,000 MRI scans, the validation set contains 12000 MRI scans, and the testing set contains 12000 MRI scans. All the transformed images in training, validation and testing are independent of each other. Table 3 illustrates the dataset’s final description after all preprocessing procedures.

TABLE 3. Final dataset after enhancement and augmentation.

Properties	Values
Total Number of Images	60000
Types of classes	6
Enhancement	CLAHE
Filtering	Green Fire Blue
NC	10000
MCI	10000
EMCI	10000
LMCI	10000
SMC	10000
AD	10000
Training Images	36000
Validation Images	12000
Testing Images	12000

IV. PROPOSED MODEL

Our proposed model was developed with the primary intent of automating the identification of persons with AD while enhancing prediction accuracy. In order to detect AD using MRI data, five pre-trained models, including the VGG16, MobileNetV2, AlexNet, ResNet50, and InceptionV3 models, are investigated to determine the most efficient transfer learning strategy for the classification job. To achieve superior classification efficiency, a modified DL framework is presented to identify AD.

A. ENVIRONMENTAL SETUP

All testing was carried out on an AMD Ryzen 7 3800X Central processing unit with an 8-core, 16-thread configuration running at 3.90 GHz and 32 GB of RAM, and the Anaconda 3 Spyder application platform was used. It is paired with a GPU from the MD Radeon RX 580 series. Following extensive analyses of five architectures, the AlzheimerNet model was selected as the most accurate model. AlzheimerNet training took over 6 days with this configuration. Python and OpenCV were used to achieve the results. The model is tested using Scikit-learn and NumPy.

B. DEEP LEARNING MODELS FOR CLASSIFICATION

1) VGG16

Simonyan and Zisserman [49] introduced VGG16 DCNN. VGG models may help the kernel learn increasingly complicated traits. A fine-tuned VGG16 model outperformed a fully trained network [46]. The network was trained on 1 million

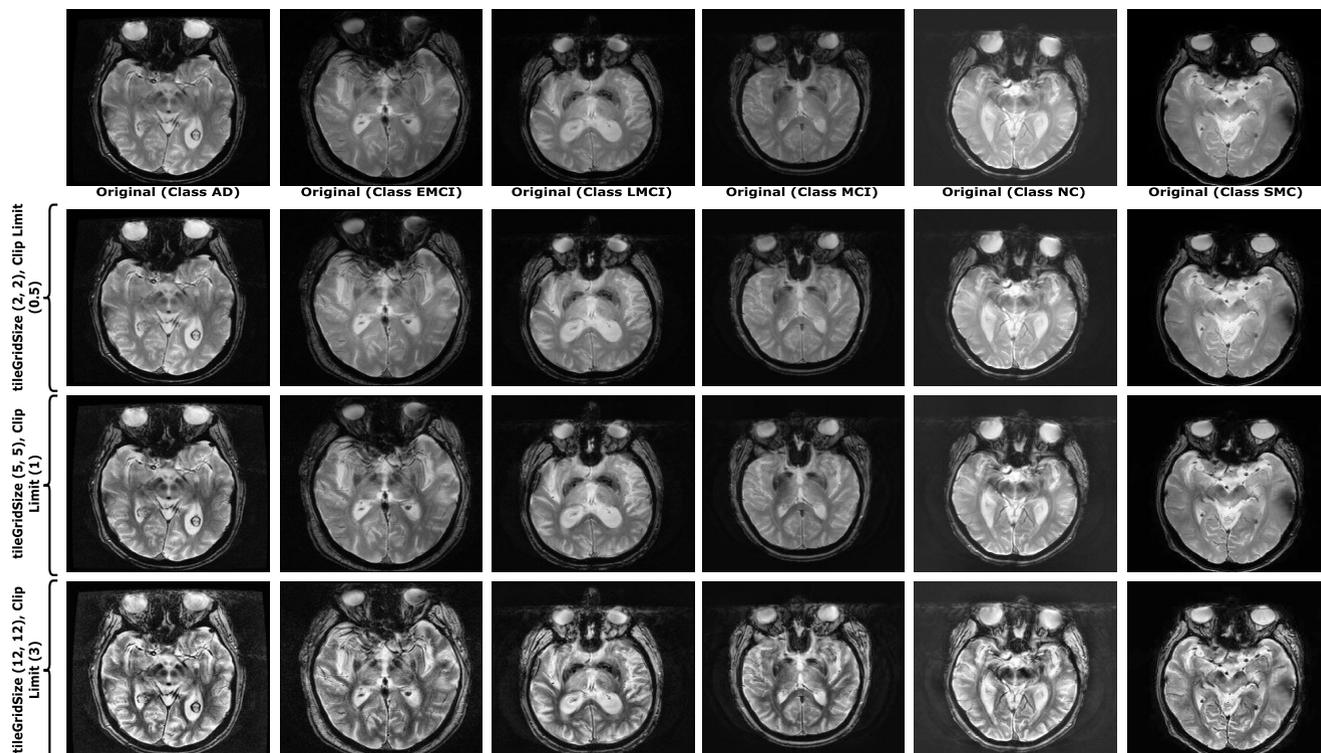


FIGURE 3. The output of CLAHE enhancement technique applied on six classes: AD, EMCI, LMCI, MCI, NC and SMC using various tileGridSize and clipLimit combinations.

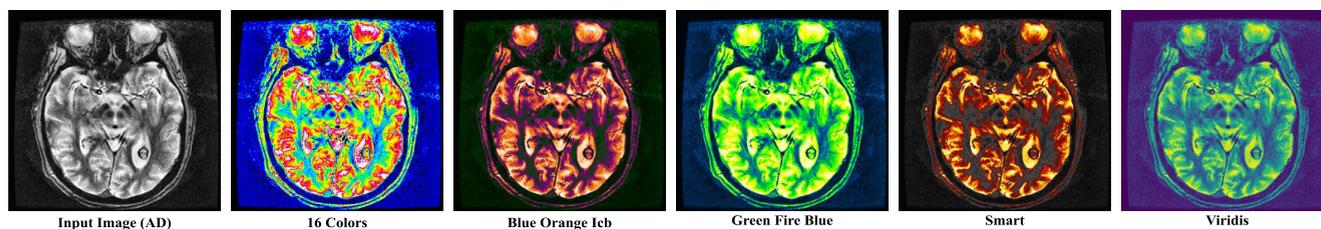


FIGURE 4. The output of various filter techniques: 16 Colors, Blue Orange Icb, Green Fire Blue, 6 Shades, Cyan Hot, applied on six classes: AD, EMCI, LMCI, MCI, NC and SMC.

ImageNet pictures. The model is made of sixteen layers and is able to identify pictures from diverse categories. The conv1 layer receives an input of 224×224 RGB images of fixed size. The image is processed through a series of convolutions with filters configured to develop the lowest achievable area of receptivity size: 224×224 . In addition, one of the combinations employs 1×1 convolution filters, which perform a linear transformation on the input channels. Following the convolutional layers, spatial pooling is achieved using five max-pooling layers. Over a 2×2 pixel frame, Stride 2 is employed to perform max-pooling. Three Convolution Layers are stacked on top of a convolutional layer stack. The first two have a total of 4096 channels, while the third is capable of 1000-way ILSVRC classification. Lastly, there is a softmax layer [50]. In all networks, the structure of the fully interconnected layers is the same [51].

2) MobileNetV2

MobileNetV2 [52] is a 53-layer convolutional neural network. It is a pre-trained classifier based on over a million images from ImageNet. The model has been pre-trained to classify images into 100 different object categories. Consequently, the network has obtained a lot of features extracted for a broad set of images. MobileNetV2’s construction starts with 32 filters and 19 bottleneck layers arranged in fully convolution layers. It consists of two separate blocks, each with three layers. Both blocks begin and conclude with 1×1 convolutional layers containing 32 filters, however, the second layer is a convolutional layer that operates in depth order. The ReLU is used at all levels of the structure. The stride values differ between the two blocks, with block one holding the stride size of 1, and block two holding the stride length of 2.

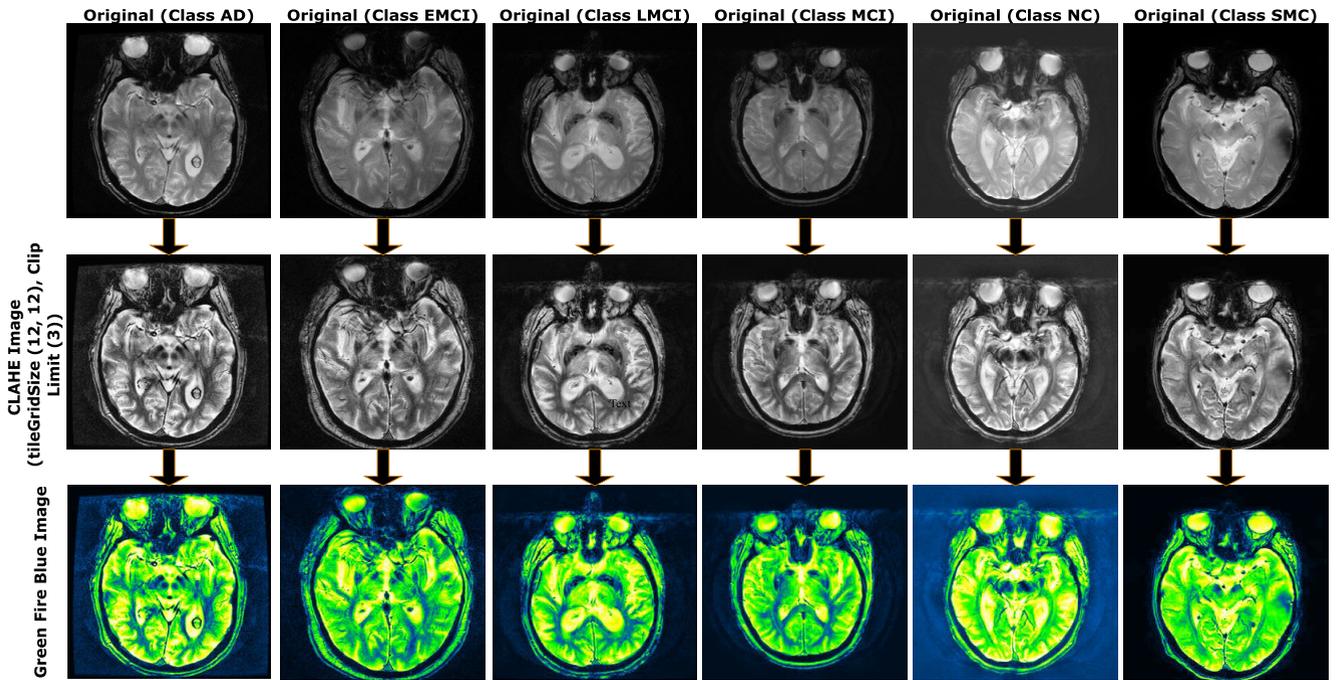


FIGURE 5. After applying image enhancement techniques, final Output of the enhanced images (CLAHE, Green Fire Blue).

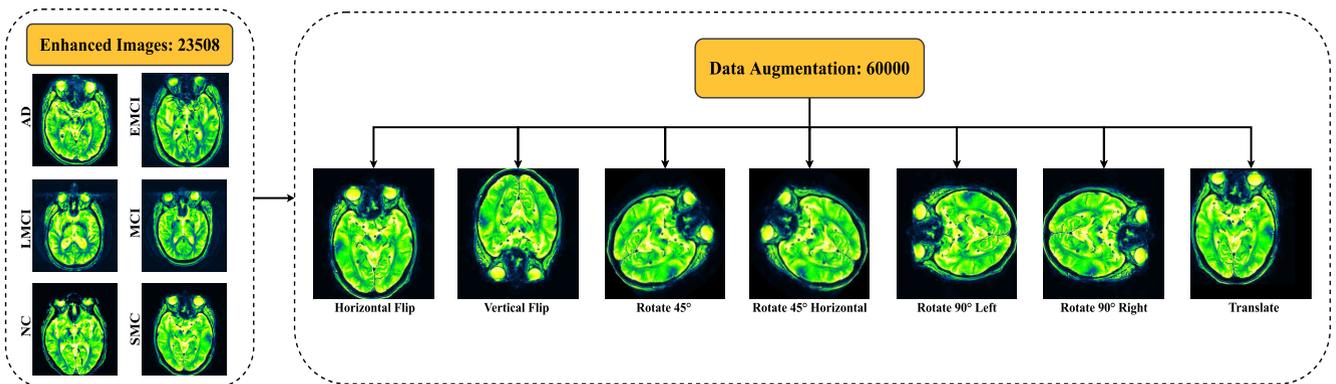


FIGURE 6. The output of applying image seven augmentation techniques: Horizontal flip, Vertical flip, Rotate 45°, Rotate 45° Horizontal, Rotate 90° left, Rotate 90° right, Translate.

3) AlexNet

AlexNet [53] is a well-known model of convolutional neural networks. As its fundamental construction blocks, it includes convolutions, max pooling, and dense layers. The model is fitted across two GPUs using grouping convolutions. Eight layers make up the Alexnet architecture, and each layer has a unique set of attributes that could be trained. The model is made up of 5 convolution layers that use a combo of max pooling and fully connected layers, as well as 2 normalizing layers and 1 softmax layer. Each layer is composed of convolution layer and a nonlinear activation function based on the ReLU principle. Max pooling is performed by the use of pooling layers. Due to the connected layers, the input size of $224 \times 224 \times 3$ is fixed. When a grayscale picture is used

as an input, it is converted to RGB by doubling the single channel, resulting in a three-channel RGB image. The model has 60 million parameters and a batch size of 128.

4) ResNet50

ResNet50 [54] is a 50-layer convolution neural network of 48 convolution layers, one Max Pool layer, and one Average Pool layer. It supports floating-point operations up to 3.8×10^9 . The ResNet50 architecture employs a combination of convolutional filters of various sizes to address the decay problem associated with CNN models and to decrease the training time associated with the deep structure. ResNets have fewer filters and are thus faster. When compared to the smaller 18-layer ResNets, the performance of the 34-layer

ResNet is 3.6 billion FLOPs, as opposed to 1.8 billion FLOPs. This architecture can be trained on approximately 23 million parameters. The network may take an input picture with a height, width, and channel width that are multiples of 32. Each ResNet design uses a 7×7 and 3×3 kernel size for initial convolution and max-pooling, respectively.

5) InceptionV3

InceptionV3 [55] minimizes the amount of computational power necessary by modifying previous Inception designs. A few techniques used to reduce computing costs include dimension reduction, regularization, factorized convolutions, and parallelized computations. This network significantly improves upon earlier Inception models in several ways, including factorized convolutional layers, label smoothing, and the inclusion of auxiliary classifiers to transmit labelled data throughout the network. The InceptionV3 model's training time is reduced by substituting smaller convolutions for bigger convolutions. Numerous optimization strategies have been proposed to remove constraints and enhance flexibility in an InceptionV3 model. Image pre-processing is a critical component of the system and significantly impacts the model's maximum accuracy during training. At the very least, photos must be encoded and scaled to match the model. For Inception, images should be $299 \times 299 \times 3$ pixels in size. The model is composed of symmetric and asymmetric foundations, including average pooling, convolutions, concatenation, maximum pooling, dropouts and fully connected layers. The model employs batch normalization considerably on activation inputs and the loss is computed using the SoftMax method.

6) PROPOSED MODEL (AlzheimerNet)

The fine-tuned InceptionV3 architecture outperforms the other five architectures mentioned above in classification performance. Thus, the AlzheimerNet model is suggested, principally on the basis of a modified InceptionV3 framework and evaluated using the ADNI dataset. Additionally, ablation research is conducted to enhance the architecture's durability for the Alzheimer's classification problem. AlzheimerNet is compact, computationally intensive, and improved performance in large and small datasets. Figure 7 depicts the model's overall structure. The dataset consists of 60,000 images into 6 classes namely NC, MCI, EMCI, LMCI, SMC, and AD. The dataset is split into 20% for testing purposes, 60% for training and 20% for validation purposes. Initially, all the images are not the same size, all the images are reshaped into 255×255 . The base layer of InceptionV3 is frozen later and different types of layers are added to customize the model for better accuracy on MRI scans dataset. The input layer of the architecture image size is $255 \times 255 \times 3$. In the first block, InceptionV3 base architecture is used and dropout was applied to avoid the overfitting problem with a range of 0.5. Convolution, GlobalAveragePooling and flatten layer is added with the kernel size 3×3 . BatchNormalization is

conducted in the second block, and a dense layer of 512 neurons with the activation function Relu is added.

Similarly, Dropout and BatchNormalization have been used in a different combination in the third block. Consequently, another dense layer with 256 neurons and a ReLU activation function has been introduced in this block. BatchNormalization and ReLU activation functions are added to simplify the process. The same method is repeated for the next two dense layers with the kernel size 3×3 , containing 128 and 64 neurons, respectively, for blocks 4 and 5.

In the last block, a flatten layer, Dropout, and BatchNormalization were applied using the RMSprop optimizer, and a softmax activation was included in the final layer. Throughout the procedure, a rate of learning of 0.00001 is utilized. Finally, the performance is evaluated based on the performance matrix. The layer's connection and novel features in the model are described stepwise below.

Input: 60000 MRI scans images (60% train, 20% validation, 20% test).

Step 1: Batch Normalization, data augmentation, preprocessing.

Step 2: Freeze the base layer and add the proposed convolution layer with image size 255×255 , optimizer='RMSProp', Activation Function='Relu', Kernel Size (3,3).

Step 3: Applying dropout(0.5), then applying GlobalAveragePooling2D.

Step 4: Flattening the image data.

Step 5: Applying BatchNormalization() along with 512 Dense layers with Relu activation function.

Step 6: Applying BatchNormalization() and dropout (0.5).

Step 7: Adding another dense layer with 256 neurons and activation function as RelU.

Step 8: Applying BatchNormalization() and dropout (0.5).

Step 9: Then Dense layer with 128 Neurons with Relu Activation function and kernel size(3,3).

Step 10: BatchNormalization with Dropout (0.5) to ignore overfitting.

Step 11: Put another dense layer with 64 neurons with Relu Activation function.

Step 12: Applying Batch Normalization, adding a flatten layer with SoftMax activation function.

Step 13: Applying RMSprop optimizer and finding accuracy, f1-score, recall, and precision.

Output: 6 types of classes (AD, CN, EMCI, LMCI, MCI, SMC).

V. ABLATION STUDY

A common technique for parameter tuning is the grid search strategy, which is employed to select the parameters, including epochs, learning rate, optimizer, dropout, and batch size. Furthermore, the ablation study has been conducted to ensure the proposed model robustness. The following elements have been affected by the ablation study: Flatten layer, Global AveragePooling2D (GAP2D), Optimizer, Loss function, Filters and learning rate.

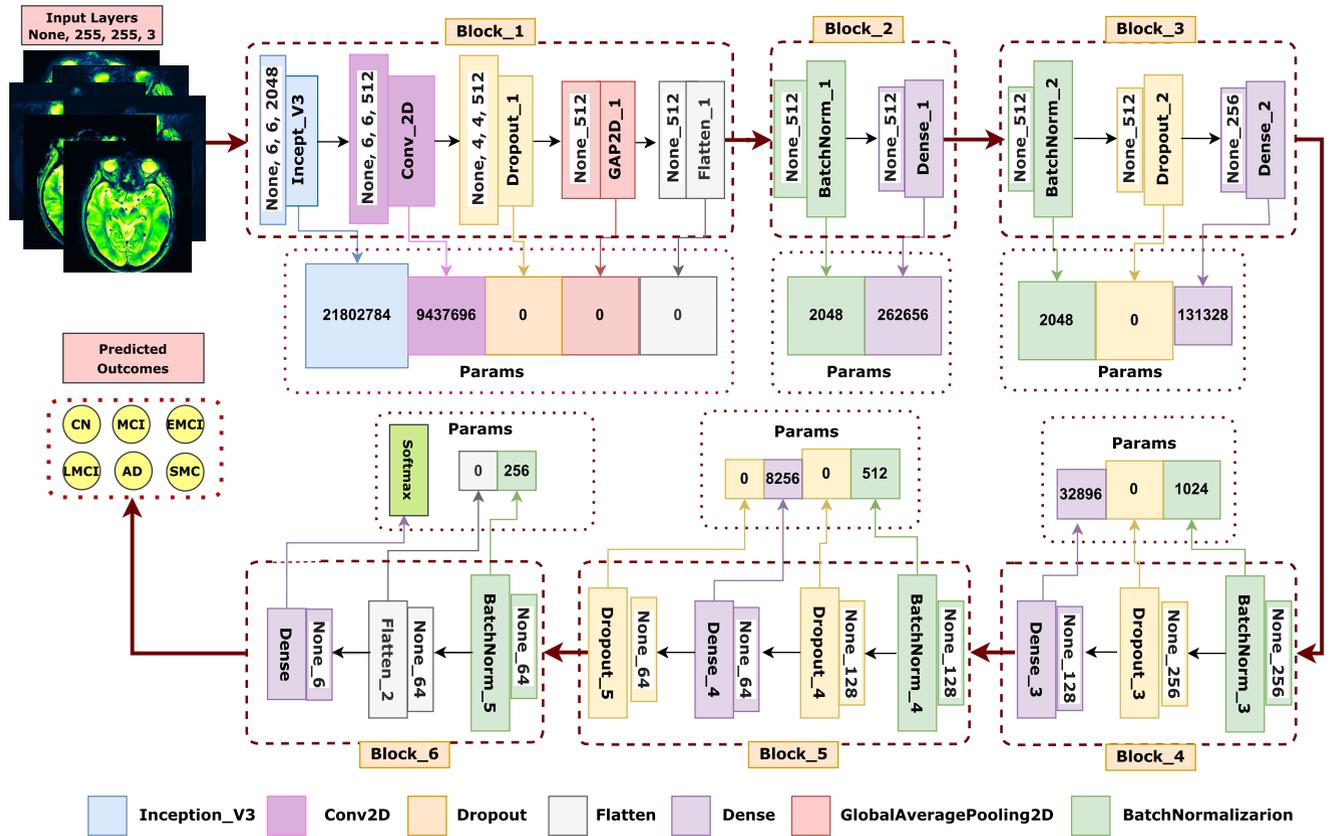


FIGURE 7. The architecture of AlzheimerNet, a fine-tuned InceptionV3.

A. CASE 1: ALTERING GLOBAL AveragePooling2D (GAP2D)

The validation loss [56] and validation accuracy are abbreviated ‘Vld_Ls’ and ‘Vld_Acc’ respectively, the test loss and test accuracy are indicated by ‘Tst_Ls’ and ‘Tst_Acc’ respectively in this research. According to Table 4, the network gives the best results for GlobalAveragePooling2D, with a Vld_Acc of 97.81% and a Tst_Acc of 98.68%. However, the accuracy for AveragePooling2D, GlobalMaxPooling2D, and Maxpooling2D reduces marginally with Vld_Acc values of 96.57%, 94.65%, 93.96% and Tst_Acc values 96.98%, 95.01%, 94.90% respectively.

B. CASE 2: ALTERING FLATTEN LAYER

In this investigation, a flatten layer has been incorporated to flatten the feature maps of the previous convolutional layers. Prior to this, a flatten layer was employed to refine the features map produced by the earlier convolutional layers. This layer is changed by GlobalAveragePooling2D, AveragePooling2D, GlobalMaxPooling2D, and Maxpooling2D to analyze its influence on the network’s performance. As seen in Table 5, the accuracy of GlobalAveragePooling2D decreases moderately. The network produces the lowest results for AveragePooling2D, GlobalMaxPooling2D, and Maxpooling2D with the Tst_Acc values 96.16%, 95.70%, and 95.10% simultaneously.

C. CASE 3: ALTERING LOSS FUNCTIONS

To decide on the proper loss function for the proposed model, examined many loss functions, including Categorical Crossentropy, Cosine similarity, and Mean Squared Error. Table 6 displays the model’s results using the specified loss functions. The model demonstrated the highest test accuracy of 98.68% when it was outfitted with Categorical Crossentropy. The Cosine similarity (95.80%) and the Mean Squared Error loss functions in fact contributed to a minor decline in test accuracy (95.01%). The categorical cross-entropy loss function was employed in subsequent studies for optimum classification performance.

D. CASE 4: ALTERING OPTIMIZERS AND LEARNING RATES

Different optimizers have been employed, including RMSprop, Adam, SGD, Nadam and Adadelta, to find the best outcome. Furthermore, the optimizer uses the momentum method to help accelerate gradients vectors in the right directions, thus leading to faster converging. Momentum was developed to enhance learning in low-curvature directions without being unstable in high-curvature directions. According to Table 8, RMSprop was the optimizer that yielded the highest performance of Tst_Acc of 98.67% and a learning rate of 0.00001. With a learning rate of 0.00001, RMSprop also had the lowest tst_Ls of 0.09. The highest Vld_Acc is 97.81%, achieved with the optimizer RMSprop and a learning

TABLE 4. Examining ablation by modifying the Global AveragePooling2D (GAP2D) layer.

Case Study	Layer Names	Vld_Acc (%)	Vld_Ls (%)	Tst_Acc (%)	Tst_Ls(%)	Performance
1	GlobalAveragePooling2D	97.81	0.10	98.68	0.09	Identical
	AveragePooling2D	96.57	0.12	96.98	0.29	Accuracy_Reduced
	GlobalMaxPooling2D	94.65	0.22	95.01	0.20	Accuracy_Reduced
	MaxPooling2D	93.96	0.34	94.90	0.31	Accuracy_Reduced

TABLE 5. Altering the Flatten layer to examine ablation study.

Case Study	Layer Names	Vld_Acc(%)	Vld_Ls (%)	Tst_Acc (%)	Tst_Ls(%)	Performance
2	Flatten	97.81	0.10	98.68	0.09	Identical
	GlobalAveragePooling2D	97.50	0.16	98.23	0.12	Identical
	AveragePooling2D	95.72	0.21	96.16	0.19	Accuracy_Reduced
	GlobalMaxPooling2D	95.56	0.28	95.70	0.22	Accuracy_Reduced
	Maxpooling2D	94.89	0.38	95.10	0.28	Accuracy_Reduced

TABLE 6. Altering the Loss function to examine ablation study.

Case Study	Loss_Functions	Vld_Acc (%)	Vld_Ls (%)	Tst_Acc (%)	Tst_Ls(%)	Performance
3	Categorical_Crossentropy	97.81	0.10	98.68	0.09	Identical
	Cosine Similarity	95.10	0.11	95.80	0.11	Accuracy_Reduced
	Mean Squared Error	94.65	0.18	95.01	0.15	Accuracy_Reduced

rate of 0.00001. The other optimizers, including Adam, SGD, Nadam, Adadelata performed well with this model, with the Tst_Acc values over 92%.

E. CASE 5: ALTERING FILTERS

A variety of filters were performed on the dataset in order to get a satisfactory result. This filtering strategy improves the image’s characteristics and makes it more comparable to the train, validate, and test models. To choose the effective filter, 5 various filters are applied to the dataset and evaluated through the model. From table 8, the ‘Green Fire Blue’ filter achieved the highest Vld_Acc 97.81% and Tst_Acc 98.68% among all the filters.

F. MODEL PARAMETERS SELECTION BASED ON ABLATION STUDY

The proposed method AlzheimerNet was shown to be more trainable on image classes than the other models. As seen in Table 9, it also produced more consistently excellent outcomes than competing models.

Following the completion of the ablation research on the proposed AlzheimerNet network, it was discovered that when the appropriate learning rate, loss function, activation function, and optimizer are applied, the model’s classification accuracy improves. Table 10 summarizes AlzheimerNet’s final configuration.

G. PERFORMANCE EVALUATION MATRIX

To evaluate the models, elements such as True-Positive (TP), False-Positive (FP), True-Negative (TN), and False-Negative

(FN) are derived from the confusion matrix produced for each model. While working with a certain class or stage of the disease, TP indicates correctly classifying data to a class. FP refers to incorrectly classifying other class data as the detecting class. TN refers to correctly identifying which class a data does not belong to. Finally, FN refers to classifying data incorrectly as another class. Applying the Confusion Matrix Elements, measures such as accuracy, precision, sensitivity, specificity, and f1-score of many others are calculated.

From Equation (6), accuracy is the number of correctly classified samples in all the classes.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

In equation (7), Accuracy_(overall) refers to the model’s overall accuracy where n is the total number of classes in the dataset.

$$Accuracy_{(overall)} = \frac{\sum TP_n}{total\ number\ of\ data} \tag{7}$$

From Equation (8), precision is the correctly classified samples from the truly positive samples of the detecting class.

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

From Equation (9), Sensitive is the number of correctly identified samples of a class from all the original samples of that class.

$$Sensitivity/Recall = \frac{TP}{TP + FN} \tag{9}$$

TABLE 7. Altering the optimizers and learning rates to examine ablation study.

Case Study	Optimizers	Learning_Rates	Vld_Acc(%)	Vld_Ls(%)	Tst_Acc(%)	Tst_Ls(%)	Performance
4	RMSprop	0.00001	97.81	0.10	98.68	0.09	Identical
		0.001	95.66	0.29	95.89	0.18	Accuracy_Reduced
		0.0001	96.95	0.16	97.50	0.10	Accuracy_Reduced
	Adam	0.01	94.72	0.57	95.01	0.31	Accuracy_Reduced
		0.00001	97.12	0.15	97.12	0.13	Accuracy_Reduced
		0.001	95.88	0.29	96.50	0.17	Accuracy_Reduced
	SGD	0.00001	96.79	0.20	97.02	0.16	Accuracy_Reduced
		0.01	94.70	0.31	95.91	0.24	Accuracy_Reduced
		0.00001	94.85	0.49	95.08	0.26	Accuracy_Reduced
	Nadam	0.001	93.62	0.62	94.02	0.51	Accuracy_Reduced
		0.0001	94.50	0.50	95.05	0.27	Accuracy_Reduced
		0.01	93.21	0.69	93.98	0.40	Accuracy_Reduced
	Adadelta	0.00001	96.50	0.16	96.97	0.11	Accuracy_Reduced
		0.001	95.52	0.16	96.01	0.17	Accuracy_Reduced
		0.0001	95.57	0.20	96.03	0.16	Accuracy_Reduced
	Adadelta	0.01	94.24	0.32	95.17	0.25	Accuracy_Reduced
		0.00001	93.98	0.31	94.80	0.23	Accuracy_Reduced
		0.001	92.46	0.60	92.97	0.39	Accuracy_Reduced
		0.0001	92.81	0.50	93.05	0.36	Accuracy_Reduced
			0.01	91.02	0.69	92.09	0.55

TABLE 8. Altering the filters to examine ablation study.

Case Study	Filters	Vld_Acc(%)	Tst_Acc(%)	Performance
5	16 Colors	92.88	93.70	Accuracy_Reduced
	Blue Orange Icb	90.21	90.98	Accuracy_Reduced
	Green Fire Blue	97.81	98.68	Identical
	6 Shades	93.70	94.91	Accuracy_Reduced
	Cyan Hot	89.12	90.29	Accuracy_Reduced

TABLE 9. Performance summary of the models.

Models	Image Mode and Overfitting	F1-Score
VGG16	MRI and No_evident (Dropout_conducted)	78.85%
MobileNetV2	MRI and No_evident (Dropout_conducted)	86.85%
AlexNet	MRI and No_evident (Dropout_conducted)	78.86%
ResNet50	MRI and No_evident (Dropout_conducted)	80.48%
InceptionV3	MRI and No_evident (Dropout_conducted)	92.32%
AlzheimerNet	MRI and No_evident (Dropout_conducted)	98.68%

From Equation (10), Specificity is the number of accurately identified samples of other classes from all the actual samples in those classes.

$$Specificity = \frac{TN}{TN + FP} \tag{10}$$

From Equation (11), F1-Score is overall the performance calculated from the precision and recall of the model.

$$F1 - Score = 2 \times \left(\frac{Precision \times Recall}{Precision + Recall} \right) \tag{11}$$

From Equation (12), Negative Predicted Value (NPV) is the number of all the samples of other classes by the models

TABLE 10. Proposed architecture configuration after ablation study.

Configurations	Values
Image Size	255×255
Epochs	300
Optimizer	RMSprop
Learning Rate	0.00001
Batch Size	200
Loss function	Categorical Crossentropy
Weight decay	0.0001
Function (activation)	Softmax
Dropout	0.5

to be of those classes.

$$NPV = \frac{TN}{TN + FN} \tag{12}$$

From Equation (13), False Negative Rate (FNR) refers to the number of predicted negative values in real positive in the data sample.

$$FNR = \frac{FN}{FN + TP} \tag{13}$$

The false Positive Rate (FPR) from Equation (14) is the number of negative samples classified as positive.

$$FPR = \frac{FP}{FP + TN} \quad (14)$$

The false Discovery Rate (FDR) from Equation (15) is the number of positive samples classified as negative.

$$FDR = \frac{FP}{FP + TP} \quad (15)$$

From Equation (16), False Omission Rate (FOR) is the number of negative values that are both predicted and real negative.

$$FOR = \frac{FN}{FN + TN} \quad (16)$$

VI. RESULTS AND DISCUSSION

A. ANALYSIS

The work employs five existing CNN models that are VGG16, MobileNetV2, AlexNet, ResNet50 and InceptionV3. The performance of the models is compared to identify the most efficient model that can classify the six classes from T2-weighted MRI scan images acquired from the ADNI database. In comparison to the other models, it is discovered that InceptionV3 is the model with the highest classification accuracy. Subsequently, the model InceptionV3 is modified to get higher classification accuracy and the fine-tuned model AlzheimerNet is proposed. The AlzheimerNet model is fine-tuned with 300 epochs and RMSprop optimizer to optimize the loss function. The learning rate of the model is 0.00001. To guarantee the highest quality data, the MRI scan images are resized to 255×255 pixels after going through a series of image preprocessing stages for improved classification accuracy. After augmentation, the balanced dataset contains a total number of 60000 MRI scan images which are divided into training, validation and test sets in the ratio of 60:20:20.

B. CONFUSION MATRIX

Figure 8 represents a visualization of the performance of each of the employed models using a confusion matrix. The classification performance of each of the classes used in the models is shown in detail. The test dataset contains a total number of 12000 Image data. Therefore, each class (a total of six classes) consists of 2000 image data in the test dataset. According to the confusion matrix, the proposed model AlzheimerNet can successfully classify 1981 images from the CN class, 1989 images from the SMC class, 1993 images from the AD class, 1954 images from MCI class, 1958 images from LMCI class, and 1966 images from EMCI class. The existing VGG16, MobileNetV2, AlexNet, ResNet50 and InceptionV3 models in total can successfully detect 9462, 10422, 9463, 9657 and 11078 images, respectively.

C. PERFORMANCE COMPARISON OF ALL APPLIED MODELS

Using equation 7, the overall accuracy, $Accuracy_{(overall)}$ of the AlzheimerNet model is achieved as 98.68%.

Similarly, the overall accuracy of existing VGG16, MobileNetV2, AlexNet, ResNet50 and InceptionV3 models are 78.85%, 86.85%, 78.86%, 80.48%, and 92.32% respectively as shown in table 11. The mean precision score, mean recall score and mean specificity score of the models are calculated by adding the values of the respective parameter of each class and dividing the sum by the number of classes. The overall f1-score of the models is determined using equation 11, where precision and recall are mean precision and mean recall. The proposed models give the best result compared to the existing models as the mean precision is 98.68%, mean recall is 98.68%, mean specificity is 99.74%, and the overall f1-score is 98.68%. In table 11, Accuracy and loss of training, validation and test are also presented apart from performance parameters as an overview of the models' complete performance where the training accuracy and loss are represented by Trn_Acc and Trn_Ls, respectively. The proposed AlzheimerNet model shows the highest accuracy during training, validation and testing with 98.95%, 97.81% and 98.68%, respectively. The computed loss function during training, validation and testing and minimum with 0.20%, 0.23% and 0.18%, respectively, which indicates that the model is more efficient compared to the existing models.

D. PERFORMANCE COMPARISON OF ALL APPLIED MODELS

From the confusion matrix depicted in figure 8, the components TP, FP, TN, and FN are computed. Accuracy, precision, recall, specificity, and f1-score are determined using these components from equations 6 to 16. The classification performance of the six classes of the dataset is visualized in detail as bar graphs. From the graphs, it can be perceived that the proposed model AlzheimerNet shows much superior performance across all parameters for the six classes. In figure 9, the accuracy (ACC), precision (PRE), recall (REC), specificity (SPE), f1-score (FSC) and NPV are scribed for the six implemented models. To compare the performance of the classification of the six classes by the implemented model, the bar diagram is presented. It is observed from figure 9(a) that, the model VGG16 achieves classification accuracy is 94.30%, 94.03%, 95.48%, 91.75%, 90.37% and 91.78% and the f1-scores are 82.35%, 81.07%, 85.18%, 76.04%, 73.70% and 76.23% for the classes NC, SMC, AD, MCI, LMCI and EMCI respectively. Even with considerably high individual classification accuracy, the overall performance cannot be considered adequate. Similarly for the classes, in figure 9(b) the model MobileNetV2 classification accuracy are 95.49%, 95.59%, 97.29%, 94.48% 94.62% and 96.23% respectively. The highest F1-score is 91.22% for the class AD and the lowest is 83.80%, for the class MCI which are not satisfactory rates as well. In figure 9(c), the accuracy scores of AlexNet are 91.93%, 92.88%, 97.16%, 89.87%, 92.73% and 93.15% respectively for the mentioned classes. The highest and lowest F1-score are 91.05% and 72.36% for the classes AD and MCI respectively. While the model is quite popular for image detection, the model failed to provide an acceptable

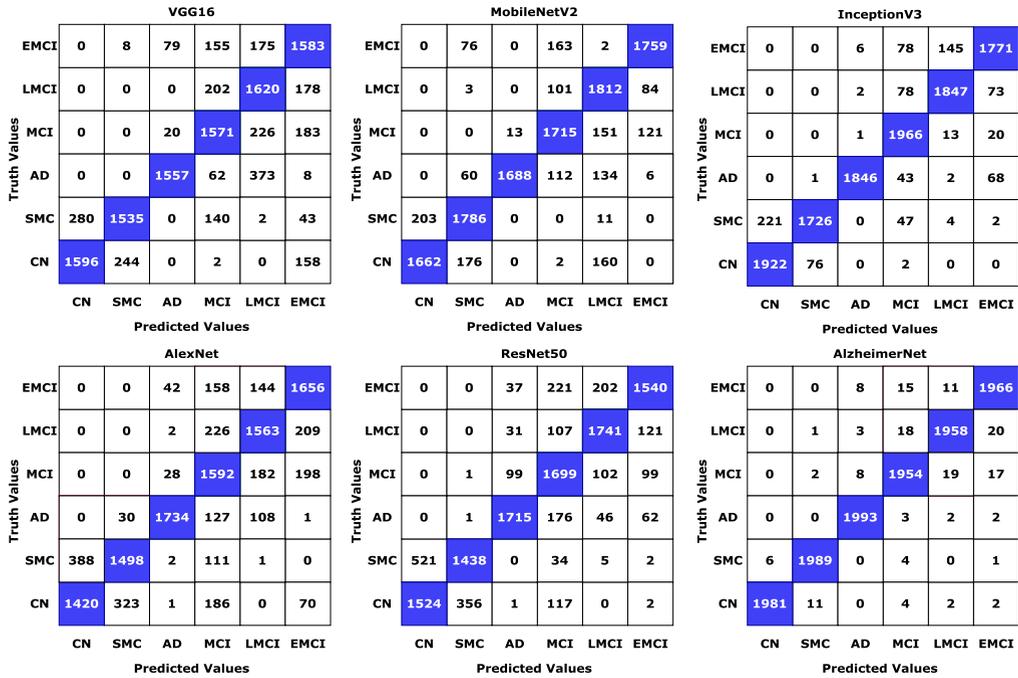


FIGURE 8. Confusion matrix of all applied models.

TABLE 11. Comparing the efficiency of the deployed transfer learning models.

Models	Overall Acc	Mean Pre	Mean Recall	Mean Specificity	Overall F1-Score	Trn_Acc %	Trn_Ls %	Vld_Acc %	Vld_Ls %	Tst_Acc %	Tst_Ls %
VGG16	78.85	79.97	78.85	95.77	79.41	79.72	0.27	78.69	0.34	78.85	0.35
MobileNetV2	86.85	87.40	86.85	97.37	87.13	88.75	0.21	86.77	0.28	86.85	0.29
AlexNet	78.86	79.58	78.86	95.77	79.22	78.91	0.41	78.12	0.30	78.86	0.28
ResNet50	80.48	80.87	80.48	96.10	80.67	81.53	0.25	80.81	0.31	80.48	0.29
InceptionV3	92.32	92.56	92.32	98.46	92.44	91.62	0.20	91.97	0.23	92.32	0.18
AlzheimerNet	98.68	98.68	98.68	99.74	98.68	98.95	0.12	97.81	0.10	98.68	0.09

classification score when applied to the ADNI dataset of MRI brain scans. In figure 9(d) the classification scores of model ResNet50 are illustrated where the model classified the classes with 91.69%, 92.33%, 96.23%, 92.03%, 94.88% and 93.78% respectively accuracy. Again the highest and lowest F1-score belongs to the class AD and NC with 88.33% and 75.35% respectively. In figure 9(e), the detailed performance of the InceptionV3 model is stated that shows the model achieves the accuracy of 97.51%, 97.08%, 98.64%, 97.65%, 97.03% and 96.73% respectively in predicting the six classes. The F1-score of the class AD is 95.77% and EMCI is 90.04%. The precision of the classes varies largely for pre-trained models. Whereas the NPV is consistently huge for all classes of the implemented models.

In figure 10, the FOR, FPR, FDR and FNR of the implemented models are stated. In figure 10(a), the highest FOR, FPR, FDR and FNR of the model VGG16 are 4.55% for SMC, 7.76% for LMCI, 32.39% for LMCI and 23.25% for SMC respectively. From figure 10(b), for the model MobileNetV2, the highest FOR score is 3.33% for the class NC. Similarly, the highest FPR, FDR and FNR scores are

4.58% for LMCI, 20.18% again for LMCI and 16.9% for NC respectively. For the model AlexNet in figure 10(c), the highest FOR, FPR, FDR and FNR scores are 5.69% for the class NC, 8.08% for MCI, 33.67% for MCI again and, 29% again for NC respectively. Figure 10(d) shows the performance of the model ResNet50, where the highest scores of FOR and FPR are 5.51% for the class SMC and 6.55% for the class MCI. Similarly the scores of FDR and FNR are 27.82% for class MCI and 28.1% for SMC. For the final pre-trained model, InceptionV3, the highest scores are 2.69% for class SMC, 2.48% for MCI, 11.2% for MCI and 13.7% for SMC of the performance measures FOR, FPR, FDR and FNR respectively.

The AlzheimerNet model proposed in the study obtains much higher performance scores compared to the five existing models. From figure 9 (f) it can be observed that the classification accuracy of the classes NC, SMC, AD, MCI, LMCI and EMCI are 99.79%, 99.79%, 99.78%, 99.25%, 99.37% and 99.37% respectively. Similarly, all classes have much higher recall, specificity, F1-score, and NPV values. The F1-score of the mentioned classes are 99.37%, 99.38%,

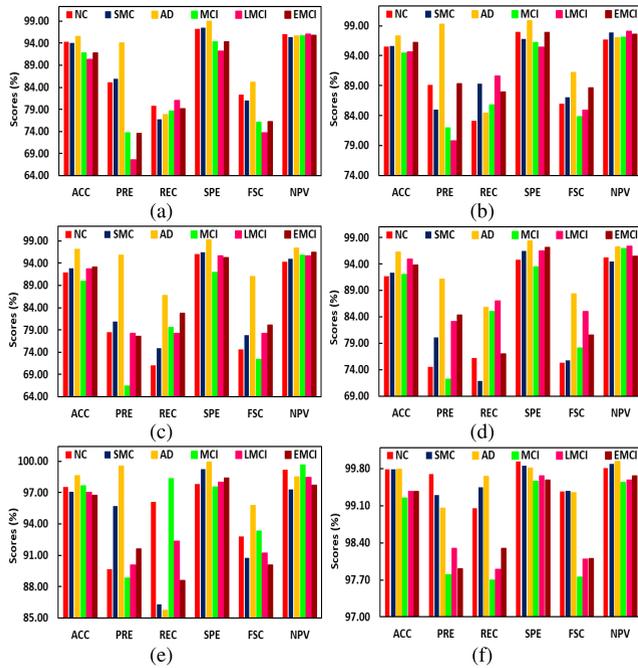


FIGURE 9. Performance scores (ACC, PRE, REC, SPE, FSC, NPV) of networks. (a) VGG16. (b) MobileNetV2. (c) AlexNet. (d) ResNet. (e) InceptionV3. (f) AlzheimerNet.

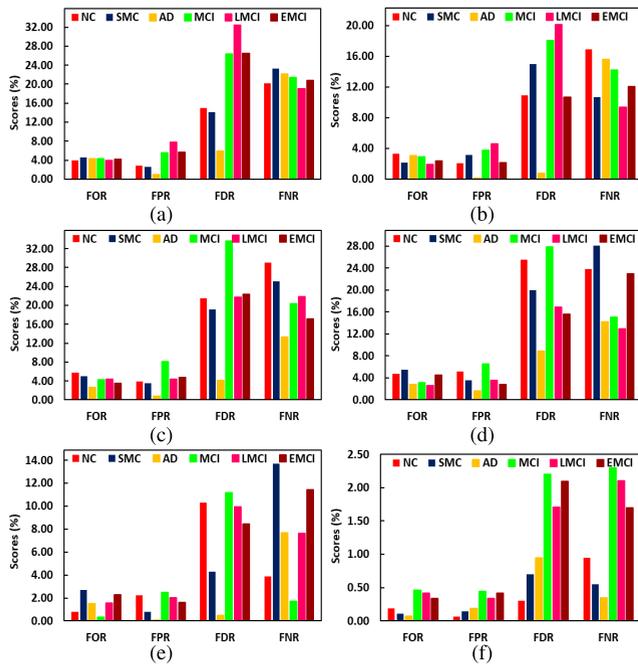


FIGURE 10. Performance scores (ACC, PRE, REC, SPE, FSC, NPV) of networks. (a) VGG16. (b) MobileNetV2. (c) AlexNet. (d) ResNet. (e) InceptionV3. (f) AlzheimerNet.

99.35%, 97.75%, 98.10% and 98.1% respectively. Furthermore from figure 10(f), the FNR, FPR, FDR and FOR of the classes are almost close to 0. The FNR and FPR scores for the classes are, 0.95%, 0.55%, 0.35%, 2.30%, 2.10% and

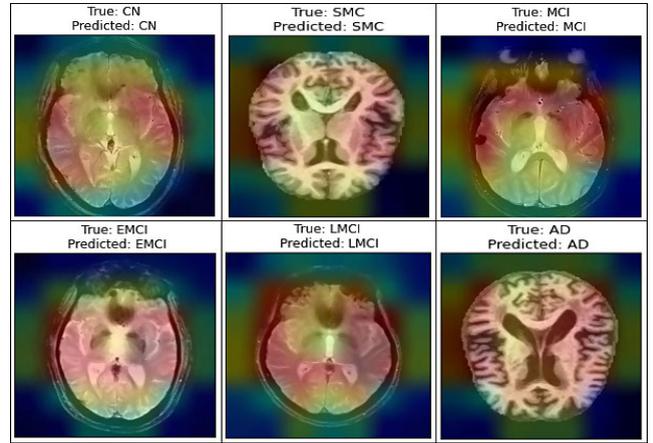


FIGURE 11. Visualization on MRI scans images of NC, SMC, MCI, EMCI, LMCI, and AD infected using Grad-CAM on the proposed model AlzheimerNet.

1.70% respectively and 0.06%, 0.14%, 0.19%, 0.44%, 0.34%, and 0.42% respectively. The highest FDR and FOR scores are 2.20% and 0.46% respectively both for the class MCI. This suggests that the model display superior performance outcome in detecting the stages of AD.

E. GRAD-CAM ANALYSIS

In order to make deep learning more understandable and applicable, numerous studies have been undertaken. Additionally, in several deep learning applications connected to medical imaging, it is crucial to increase the comprehensibility of deep neural networks. Gradient Weighted Class Activation Mapping (Grad-CAM), a method created by Selvaraju et al. [57], serves as an example of how deep learning works. A visual depiction of any densely connected neural network is provided by Grad-CAM. This helps when making predictions or doing recognition tasks to learn more details about the model. The proposed model is employed as a detection method, with the input being an image from a T2-weighted MRI scan. The final Convolution layer receives Grad-CAM after computing the anticipated label using the suggested model. Here, the Grad-CAM approach is utilized to extract the feature map for the proposed network. As a visual representation of a proposed network, the heat map displays the image area crucial for identifying the target class. The proposed methodology for visualising heatmaps on MRI scan pictures is shown in Figure 11. The figure depicts the heatmaps of the original images of a CN, SMC, MCI, EMCI, LMCI, and AD that were generated by utilizing the Grad-CAM algorithm. This highlights the areas of the brain that have the most impact on the prediction of the model. In order to validate the probable diagnosis, this information may be significant in directing Alzheimer’s detection. Additionally, regular application of the algorithm with integrated visual characteristics to various patients would raise a doctor’s confidence in the algorithm’s predictions if the system is correct in its predictions.

F. ROC CURVE ANALYSIS

To measure the suggested model’s performance, the ROC (Receiver Operating Characteristics Curve) and AUC (Area Under Curve) [58] values are computed. For multiclass classification, one class vs rest process is applied. The ROC curve is constructed with sensitivity (True Positive rate) as the y-axis and 1-Specificity (false positive rate) as the x-axis. The AUC score is determined by calculating the area under the ROC curve. The limit of the AUC value is 0 to 1. The closer the value is to 0, the poor the performance of the model is. Similarly, the closer the value is to 1, the better the performance of the model is. In figure 12 the ROC curve of the six classes is presented for the proposed AlzheimerNet model. The AUC value of the class CN is 0.9773, SMC is 0.9785, AD is 0.9796, MCI is 0.9733, LMCI is 0.9741, and EMCI is 0.9758. Observing the values, it can be stated that the proposed model is an acceptable model.

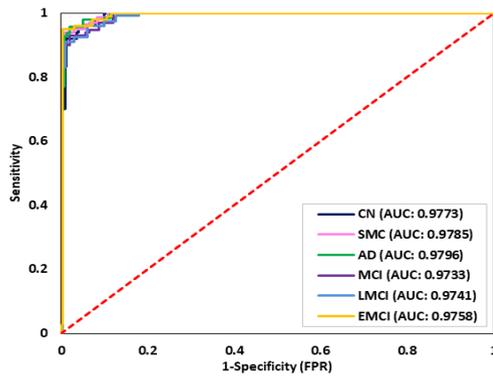


FIGURE 12. ROC curve and AUC value of the AlzheimerNet model.

G. WILCOXON SIGNED-RANK TEST

A statistical analysis of significance (S) was carried out to ensure that the results could not have been the result of random chance. The researchers accomplished this by doing the Wilcoxon signed-rank tests and computing the p-values for each model. The Wilcoxon signed-rank test is often used to compare two non-parametrically. This test was used to compare two independent samples in order to do a pairwise differences assessment on multiple observations on a single piece. Following that, the outcome indicates whether or not their population mean ranks differ. Table 12 summarizes the p-values for pairwise comparisons of the models [59], [60]. AlzheimerNet outperformed the other models. To describe, the p-value difference between AlzheimerNet and the other models is less than 0.05, indicating that AlzheimerNet outperformed the other five models statistically significantly.

H. COMPARISON ACCURACY WITH SOME EXISTING STUDIES

This article provides a comparison for determining a CNN model for identifying stages of Alzheimer’s disease.

TABLE 12. Wilcoxon signed-rank test.

No.	Pairwise_Model_Comparisons	p Value	S < 0.05
1	AlzheimerNet versus VGG16	0.0067	Pass
2	AlzheimerNet versus MobileNetV2	0.0001	Pass
3	AlzheimerNet versus AlexNet	0.0068	Pass
4	AlzheimerNet versus ResNet50	0.0030	Pass
5	AlzheimerNet versus InceptionV3	0.0021	Pass

A comparison of the results obtained using the employed pre-trained models with the results obtained using the suggested modified model demonstrates that the suggested model yields a higher performance score. In order to determine how accurate the recommended model is, it was compared with other models that have been utilized in previous research, as can be seen in table 13. The works of literature have been chosen in accordance with the AD detection models. The table explicitly demonstrates that the suggested model has far higher classification accuracy than previous works.

TABLE 13. Comparative analysis of the proposed method with some existing studies.

Authors	Class	Methods	Accuracy
Liu et al. [12]	3	3D CNN	93.26%
Amoroso et al. [13]	4	RF, DNN, Fuzzy Logic	38.8%
Yang et al. [14]	2	3DRESNET	79.4%
Khvostikov et al. [15]	3	CNN	96.7%
Lin et al. [16]	3	CNN	79.9%
Wee et al. [17]	2	Graph-CNN	89.4%
Islam et al. [18]	3	CNN	88.76%
Ahmed et al. [19]	3	CNN	90.05%
Parmar et al. [20]	2	3D CNN	96.55%
Duc et al. [21]	2	3D CNN and SVM-RFE	85.27%
Li et al. [22]	3	3D CNN and C3D-LSTM	89.47%
Parmar et al. [23]	4	Modified 3D CNN	93%
Amini et al. [24]	4	CNN	96.7%
Ebrahimi et al. [25]	2	2D CNN + RNN	96.88%
Murugan et al. [26]	5	DEMNET	84.83%
Dyrba et al. [27]	2	3D CNN	94.9%
Fu’adah et al. [28]	4	AlexNet	95%
Angkoso et al. [29]	3	Multiplane CNN	93%
Heising et al. [30]	2	2D CNN	84%
Oktavian [31]	3	ResNet18	88.3%
Yee et al. [32]	2	3D CNN	88%
Shamrat et al.	6	Proposed Model	98.68%

VII. CONCLUSION

Alzheimer’s disease is a detrimental form of dementia that affects a sizable percentage of the world’s geriatric population. AD is a degenerative neurological disorder that typically affects the elderly. It is responsible for various diseases, including memory loss and confusion. Individuals who have AD have the inability to do everyday tasks. It cannot be emphasized how critical it is to detect this condition early to provide appropriate therapy and improve patients’ lives. In this study, a fine-tuned model AlzheimerNet is proposed to classify the stages of Alzheimer’s disease: AD, EMCI,

LMCI, MCI, SMC, and NC. The raw data from an MRI scan are given a quality boost with the use of the image enhancement method. For classification, five pre-trained models and one fine-tuned model are experimented with to evaluate the performance with the highest accuracy. Different layers and hyperparameters were used to create AlzheimerNet's architecture, which was inspired by the InceptionV3 model. An ablation study on the proposed network is done to see how well it stands up with various hyperparameters. The enhanced image dataset was increased by seven augmentation methods to enlarge and balance the MRI dataset into 60000 images. AlzheimerNet performed best with RMSprop optimizer and learning rate of 0.00001, with the training accuracy of 98.95%, validation accuracy of 97.81% and testing accuracy of 98.68%. There was a high rate of correct classification with the model that used image processing, fine-tuning, and ablation study. To evaluate the proposed model, two-tailed Wilcoxon signed-rank test is performed. Results from this research show that image processing and enhancement methods may boost the model's accuracy. Transfer learning can be an excellent way to work with both small and large numbers of images in computer vision. Additionally, how the model's architecture and hyper-parameters are set up using an ablation study also has an effect on how accurate it is. As future work, a hybrid model could be implemented on ADNI fMRI and PET datasets to diagnose the stages of Alzheimer's disease.

REFERENCES

- [1] K. M. Langa, "Is the risk of Alzheimer's disease and dementia declining?" *Alzheimer's Res. Therapy*, vol. 7, no. 1, pp. 1–4, 2015.
- [2] L. E. Hebert, J. Weuve, P. A. Scherr, and D. A. Evans, "Alzheimer disease in the United States (2010–2050) estimated using the 2010 census," *Neurology*, vol. 80, no. 19, pp. 1778–1783, May 2013.
- [3] D. van Vliet, M. E. de Vugt, C. Bakker, Y. A. L. Pijnenburg, M. J. F. J. Vernooij-Dassen, R. T. C. M. Koopmans, and F. R. J. Verhey, "Time to diagnosis in young-onset dementia as compared with late-onset dementia," *Psychol. Med.*, vol. 43, no. 2, pp. 423–432, Feb. 2013.
- [4] I. O. Korolev, "Alzheimer's disease: A clinical and basic science review," *Med. Student Res. J.*, vol. 4, no. 1, pp. 24–33, 2014.
- [5] D. Mehta, R. Jackson, G. Paul, J. Shi, and M. Sabbagh, "Why do trials for Alzheimer's disease drugs keep failing? A discontinued drug perspective for 2010–2015," *Exp. Opinion Investigational Drugs*, vol. 26, no. 6, pp. 735–739, Jun. 2017.
- [6] M. Prince, R. Bryce, and C. Ferri, *World Alzheimer Report 2011: The Benefits of Early Diagnosis and Intervention*. London, U.K.: Alzheimer's Disease International, 2011.
- [7] G. M. McKhann, D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. Jack Jr., C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, R. Mayeux, R. C. Mohs, J. C. Morris, M. N. Rossor, P. Scheltens, M. C. Carrillo, B. Thies, S. Weintraub, and C. H. Phelps, "The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's Dementia*, vol. 7, no. 3, pp. 263–269, 2011.
- [8] R. A. Sperling, P. S. Aisen, L. A. Beckett, D. A. Bennett, S. Craft, A. M. Fagan, T. Iwatsubo, C. R. Jack Jr., and J. Kaye, "Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's Dementia*, vol. 7, no. 3, pp. 280–292, May 2011.
- [9] A. Wimo, L. Jönsson, J. Bond, M. Prince, and B. Winblad, "The worldwide economic impact of dementia 2010," *Alzheimer's Dementia*, vol. 9, no. 1, pp. 1–11, 2013.
- [10] (2019). *Alzheimer's Disease Fact Sheet*. Accessed: Jan. 5, 2022. [Online]. Available: <https://www.nia.nih.gov/health/alzheimers-disease-fact-sheet>
- [11] *Alzheimer's Disease Neuroimaging Initiative (ADNI) Database*. Accessed: Jan. 5, 2022. [Online]. Available: <http://adni.loni.usc.edu/>
- [12] M. Liu, D. Cheng, K. Wang, and Y. Wang, "Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis," *Neuroinformatics*, vol. 16, nos. 3–4, pp. 295–308, Oct. 2018.
- [13] N. Amoroso, D. Diacono, A. Fanizzi, M. La Rocca, A. Monaco, A. Lombardi, C. Guaragnella, R. Bellotti, and S. Tangaro, "Deep learning reveals Alzheimer's disease onset in MCI subjects: Results from an international challenge," *J. Neurosci. Methods*, vol. 302, pp. 3–9, May 2018.
- [14] C. Yang, A. Rangarajan, and S. Ranka, "Visual explanations from deep 3D convolutional neural networks for Alzheimer's disease classification," in *Proc. AMIA Annu. Symp.*, 2018, p. 1571.
- [15] A. Khvostikov, K. Aderghal, J. Benois-Pineau, A. Krylov, and G. Catheline, "3D CNN-based classification using sMRI and MD-DTI images for Alzheimer disease studies," 2018, *arXiv:1801.05968*.
- [16] W. Lin, T. Tong, Q. Gao, D. Guo, X. Du, Y. Yang, G. Guo, M. Xiao, M. Du, and X. Qu, "Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment," *Frontiers Neurosci.*, vol. 12, p. 777, Nov. 2018.
- [17] C.-Y. Wee, C. Liu, A. Lee, J. S. Poh, H. Ji, and A. Qiu, "Cortical graph neural network for AD and MCI diagnosis and transfer learning across populations," *NeuroImage. Clin.*, vol. 23, Jul. 2019, Art. no. 101929.
- [18] J. Islam and Y. Zhang, "Understanding 3D CNN behavior for Alzheimer's disease diagnosis from brain PET scan," 2019, *arXiv:1912.04563*.
- [19] S. Ahmed, K. Y. Choi, J. J. Lee, B. C. Kim, G.-R. Kwon, K. H. Lee, and H. Y. Jung, "Ensembles of patch-based classifiers for diagnosis of Alzheimer diseases," *IEEE Access*, vol. 7, pp. 73373–73383, 2019.
- [20] H. S. Parmar, B. Nutter, R. Long, S. Antani, and S. Mitra, "Deep learning of volumetric 3D CNN for fMRI in Alzheimer's disease classification," in *Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging*, vol. 11317. USA: SPIE, 2020, pp. 66–71.
- [21] N. T. Duc, S. Ryu, M. N. I. Qureshi, M. Choi, K. H. Lee, and B. Lee, "3D-deep learning based automatic diagnosis of Alzheimer's disease with joint MMSE prediction using resting-state fMRI," *Neuroinformatics*, vol. 18, no. 1, pp. 71–86, Jan. 2020.
- [22] W. Li, X. Lin, and X. Chen, "Detecting Alzheimer's disease based on 4D fMRI: An exploration under deep learning framework," *Neurocomputing*, vol. 388, pp. 280–287, May 2020.
- [23] H. Parmar, B. Nutter, R. Long, S. Antani, and S. Mitra, "Spatiotemporal feature extraction and classification of Alzheimer's disease using deep learning 3D-CNN for fMRI data," *J. Med. Imag.*, vol. 7, no. 5, p. 56001, Oct. 2020.
- [24] M. Amini, M. Pedram, A. Moradi, and M. Ouchani, "Diagnosis of Alzheimer's disease severity with fMRI images using robust multitask feature extraction method and convolutional neural network (CNN)," *Comput. Math. Methods Med.*, vol. 2021, pp. 1–15, Apr. 2021.
- [25] A. Ebrahimi and S. Luo, "Convolutional neural networks for Alzheimer's disease detection on MRI images," *J. Med. Imag.*, vol. 8, no. 2, 2021, Art. no. 24503.
- [26] S. Murugan, C. Venkatesan, M. G. Sumithra, X.-Z. Gao, B. Elakkiya, M. Akila, and S. Manoharan, "DEMNET: A deep learning model for early diagnosis of Alzheimer diseases and dementia from MR images," *IEEE Access*, vol. 9, pp. 90319–90329, 2021.
- [27] M. Dyrba, M. Hanzig, S. Altenstein, S. Bader, T. Ballarini, F. Brosseron, K. Buerger, D. Cantré, P. Dechent, L. Dobisch, E. Dützel, M. Ewers, K. Fliessbach, W. Glanz, and J.-D. Haynes, "Improving 3D convolutional neural network comprehensibility via interactive visualization of relevance maps: Evaluation in Alzheimer's disease," *Alzheimer's Res. Therapy*, vol. 13, no. 1, p. 191, 2021.
- [28] Y. N. Fu'adah, I. Wijayanto, N. K. C. Pratiwi, F. F. Taliningsih, S. Rizal, and M. A. Pramudito, "Automated classification of Alzheimer's disease based on MRI image processing using convolutional neural network (CNN) with AlexNet architecture," *J. Phys., Conf.*, vol. 1844, no. 1, Mar. 2021, Art. no. 012020.
- [29] C. V. Angkoso, H. P. A. Tjahyaningtjas, M. H. Purnomo, and I. K. E. Purnama, "Multiplane convolutional neural network (Mp-CNN) for Alzheimer's disease classification," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 1, pp. 329–340, 2022.
- [30] L. Heising and S. Angelopoulos, "Operationalising fairness in medical AI adoption: Detection of early Alzheimer's disease with 2D CNN," *BMJ Health Care Inf.*, vol. 29, no. 1, Apr. 2022, Art. no. e100485.

- [31] M. W. Oktavian, N. Yudistira, and A. Ridok, "Classification of Alzheimer's disease using the convolutional neural network (CNN) with transfer learning and weighted loss," 2022, *arXiv:2207.01584*.
- [32] E. Yee, D. Ma, K. Popuri, L. Wang, and M. F. Beg, "Construction of MRI-based Alzheimer's disease score based on efficient 3D convolutional neural network: Comprehensive validation on 7,902 images from a multi-center dataset," *J. Alzheimer's Disease*, vol. 79, no. 1, pp. 47–58, Jan. 2021.
- [33] M. Odusami, R. Maskeliunas, and R. Damasevicius, "An intelligent system for early recognition of Alzheimer's disease using neuroimaging," *Sensors*, vol. 22, no. 3, p. 740, Jan. 2022.
- [34] M. Odusami, R. Maskeliunas, R. Damasevicius, and S. Misra, "ResD hybrid model based on ResNet18 and DenseNet121 for early Alzheimer disease classification," in *Proc. 21st Int. Conf. Intell. Syst. Design Appl. (ISDA)*, Dec. 2021, pp. 296–305.
- [35] D. AlSaeed and S. F. Omar, "Brain MRI analysis for Alzheimer's disease diagnosis using CNN-based feature extraction and machine learning," *Sensors*, vol. 22, no. 8, p. 2911, Apr. 2022.
- [36] M. Odusami, R. Maskeliunas, R. Damasevicius, and S. Misra, "Comparable study of pre-trained model on Alzheimer disease classification," in *Proc. 21st Int. Conf. Comput. Sci. Appl.*, Cagliari, Italy, Sep. 2021, pp. 63–74.
- [37] N. Deepa and S. P. Chokkalingam, "Optimization of VGG16 utilizing the arithmetic optimization algorithm for early detection of Alzheimer's disease," *Biomed. Signal Process. Control*, vol. 74, Apr. 2022, Art. no. 103455.
- [38] P. Pandey, A. Khare, and P. Srivastava, "Detection of Alzheimer's disease using CNN architectures," *ADB U. J. Eng. Technol.*, vol. 11, no. 1, pp. 1–6, 2022.
- [39] A. Bhagat, S. I. Ansarullah, M. T. B. Othman, Y. Hamid, H. K. Alkahtani, I. Ullah, and H. Hamam, "A novel framework for classification of different Alzheimer's disease stages using CNN model," *Electronics*, vol. 12, no. 2, p. 469, 2023.
- [40] X. Xing, M. U. Rafique, G. Liang, H. Blanton, Y. Zhang, C. Wang, N. Jacobs, and A.-L. Lin, "Efficient training on Alzheimer's disease diagnosis with learnable weighted pooling for 3D PET brain image classification," *Electronics*, vol. 12, no. 2, p. 467, Jan. 2023.
- [41] W. S. Gan, *Signal Processing and Image Processing for Acoustical Imaging*. Cham, Switzerland: Springer, 2020.
- [42] T. Mahmood, J. Li, Y. Pei, and F. Akhtar, "An automated in-depth feature learning algorithm for breast abnormality prognosis and robust characterization from mammography images using deep transfer learning," *Biology*, vol. 10, no. 9, p. 859, Sep. 2021.
- [43] P. Wang, J. Wang, Y. Li, P. Li, L. Li, and M. Jiang, "Automatic classification of breast cancer histopathological images based on deep feature fusion and enhanced routing," *Biomed. Signal Process. Control*, vol. 65, Mar. 2021, Art. no. 102341.
- [44] N. Hassan, S. Ullah, N. Bhatti, H. Mahmood, and M. Zia, "The retinex based improved underwater image enhancement," *Multimedia Tools Appl.*, vol. 80, no. 2, pp. 1839–1857, Jan. 2021.
- [45] S. Montaha, S. Azam, A. K. M. R. H. Rafid, P. Ghosh, M. Hasan, M. Jonkman, and F. D. Boer, "BreastNet18: A high accuracy fine-tuned VGG16 model evaluated using ablation study for diagnosing breast cancer from enhanced mammography images," *Biology*, vol. 10, no. 12, p. 1347, 2021.
- [46] L. T. Menon, I. A. Laurensi, M. C. Penna, L. E. S. Oliveira, and A. S. Britto, "Data augmentation and transfer learning applied to charcoal image classification," in *Proc. Int. Conf. Syst., Signals Image Process. (IWSSIP)*, Jun. 2019, pp. 69–74.
- [47] L. G. Falconi, M. Perez, W. G. Aguila, and A. Conci, "Transfer learning and fine tuning in breast mammogram abnormalities classification on CBIS-DDSM database," *Adv. Sci., Technol. Eng. Syst. J.*, vol. 5, no. 2, pp. 154–165, 2020.
- [48] S. Akter, F. M. J. M. Shamrat, S. Chakraborty, A. Karim, and S. Azam, "COVID-19 detection using deep learning algorithm on chest X-ray images," *Biology*, vol. 10, no. 11, p. 1174, Nov. 2021.
- [49] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [50] M. F. A. Foyzal, M. S. Islam, A. Karim, and N. Neehal, "Shot-Net: A convolutional neural network for classifying different cricket shots," in *Proc. Int. Conf. Recent Trends Image Process. Pattern Recognit.*, Solapur, India, Dec. 2019, pp. 111–120.
- [51] V. R. Manaswi and B. Sankarababu, "A flexible accession on brain tumour detection and classification using VGG16 model," in *Smart Intelligent Computing and Applications*, vol. 1. Singapore: Springer, 2022, pp. 225–238.
- [52] T. H. Arfan, M. Hayaty, and A. Hadinegoro, "Classification of brain tumours types based on MRI images using mobilenet," in *Proc. 2nd Int. Conf. Innov. Creative Inf. Technol. (ICITech)*, Sep. 2021, pp. 69–73.
- [53] L. S. Kumar, S. Hariharasitaraman, K. Narayanasamy, K. Thinakaran, J. Mahalakshmi, and V. Pandimurugan, "AlexNet approach for early stage Alzheimer's disease detection from MRI brain images," *Mater. Today, Proc.*, vol. 51, no. 1, pp. 58–65, 2022.
- [54] L. Fulton, D. Dolezel, J. Harrop, Y. Yan, and C. Fulton, "Classification of Alzheimer's disease with and without imagery using gradient boosted machines and ResNet-50," *Brain Sci.*, vol. 9, no. 9, p. 212, Aug. 2019.
- [55] J. Li, P. Wang, Y. Li, Y. Zhou, X. Liu, and K. Luan, "Transfer learning of pre-trained inception-V3 model for colorectal cancer lymph node metastasis classification," in *Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA)*, Aug. 2018, pp. 1650–1654.
- [56] M. S. Junayed, A. A. Jeny, S. T. Atik, N. Neehal, A. Karim, S. Azam, and B. Shanmugam, "AcneNet—A deep CNN based classification approach for acne classes," in *Proc. 12th Int. Conf. Inf. Commun. Technol. Syst. (ICTS)*, Jul. 2019, pp. 203–208.
- [57] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2017, pp. 618–626.
- [58] X. Zhang, X. Li, Y. Feng, and Z. Liu, "The use of ROC and AUC in the validation of objective image fusion evaluation metrics," *Signal Process.*, vol. 115, pp. 38–48, Oct. 2015.
- [59] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Hoboken, NJ, USA: Wiley, 2014.
- [60] C. Nadeau and Y. Bengio, "Inference for the generalization error," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 12, 1999, pp. 1–7.



F M JAVED MEHEDI SHAMRAT received the B.Sc. degree in software engineering from Daffodil International University, in 2018. He was formerly employed with Daffodil International University and also worked as a Lecturer with the Department of Computer Science and Engineering, European University of Bangladesh. He has been actively engaged in collaborative research with researchers from Bangladesh, USA, Canada, China, and Australia. He has several research publications published in prestigious journals and conferences (Scopus). His research interests include the study include the IoT, deep learning, image processing, neural networks, bioinformatics, and machine learning.



SHAMIMA AKTER received the B.S. degree in biochemistry and molecular biology from the University of Dhaka, Bangladesh, and the M.S. degree in genomics and bioinformatics from Virginia Tech, in 2017, with a focus on machine learning research. She is currently pursuing the Ph.D. degree in bioinformatics and computational biology with George Mason University. From 2017 to 2019, she was a Research Associate in biomedical science with Virginia Tech to perform bioinformatics that includes Python, R, and Linux programming-based complex data analysis and machine learning-based virology/microbiology projects. She also performed networks analysis, different machine learning projects both in Python and R, and some deep learning projects in Python. Her research interests include artificial intelligence and machine learning in biomedical science i.e., liver disease, chronic kidney disease, mathematical modeling of Alzheimer's diseases, and computational biology.



SAMI AZAM is currently a leading Researcher and a Senior Lecturer with the College of Engineering and IT, Charles Darwin University, Casuarina, NT, Australia. He has several publications in peer-reviewed journals and international conference proceedings. His research interests include computer vision, signal processing, artificial intelligence, and biomedical engineering.



KHAN MD. HASIB (Member, IEEE) received the B.Sc. degree from the Computer Science and Engineering Department, Ahsanullah University of Science and Technology (AUST), and the M.Sc. degree from the Computer Science and Engineering Department, BRAC University. He has more than four years of teaching and three years of research experiences in computer science. He has been heavily involved in collaborative research activities especially in the fields of applied machine learning, medical image processing, health informatics, computer vision, and natural language processing. He has published over 25 research papers in highly recognized journals, book chapters, and conference proceedings. He is currently working on several projects, such as efficient detection of specific language impairment in children, applied deep learning in image-based cervical cancer detection, activity recognition using LFR, feature-based CT image registration of liver cancer, and COVID-19 vaccination prediction from numeric data using machine learning algorithms.



ASIF KARIM (Member, IEEE) is currently a Research Active Lecturer with Charles Darwin University, Australia. Alongside being an Active Researcher, he has considerable industry experience in IT, primarily in the field of software engineering. His research interests include machine intelligence, health informatics, and smart contracts.



FRISO DE BOER is currently a Professor with the College of Engineering, IT and Environment, Charles Darwin University, Australia. His research interests include signal processing, biomedical engineering, and mechatronics.



deep learning, image processing, computer vision, health informatics, smart vehicular transportation, and wireless communications.

PRONAB GHOSH received the Bachelor of Science (B.Sc.) degree from the Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh, and the master's degree from the Department of Computer Science (Artificial Intelligence), Lakehead University, ON, Canada. His several research works have been published at conferences and in reputable journals. His current research interests include artificial intelligence, machine learning,



KAWSAR AHMED (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. (Engineering) degrees in information and communication technology (ICT) from Mawlana Bhashani Science and Technology University, Tangail, Bangladesh. He is currently pursuing the Ph.D. degree in biomedical engineering with the University of Saskatchewan, Canada. He is also working as an Assistant Professor with Mawlana Bhashani Science and Technology University. Prior to that, he joined the Software Engineering Department, Daffodil International University, as a Lecturer. He has more than 250 publications in IEEE, IET, OSA, Elsevier, Springer, ISI, and PubMed-indexed journals. He has published two books on bioinformatics and photonic sensor design. He is also a Research Coordinator of the Group of Biophotomati. His research interests include biomedical engineering, biophotonics, biosensor, machine learning, federated learning, data mining, and bioinformatics. He is also a member of SPIE and OSA. He holds the top position at his department and in university and is listed among the top ten researchers in Bangladesh, from 2017 to 2020 (Scopus indexed-based). His research group received the SPIE Travelling Award and the Best Paper Award at the IEEE WIECON ECE-2015 Conference. He has achieved gold medals for engineering faculty first both in the B.Sc. and M.Sc. degrees from Mawlana Bhashani Science and Technology University for his academic excellence.



ZARRIN TASNIM received the bachelor's degree in software engineering from Daffodil International University, Dhaka, Bangladesh. Several of her research papers have been published in recognized journals and conferences. Her research interests include artificial intelligence, health informatics, and deep learning.

...