Alzheimer's Disease Detection Utilizing Pretrained Deep Learning Models on Structural MRI Scans

Akshay Kumar Sagar
Computer Science Engineering
Sagar Institute of Research and Technology
Bhopal

Dr. Ritu Shrivastava
Prof and Head
Sagar Institute of Research and Technology
Bhopal

Abstract— **Diseases affecting the brain rank high on some of the most challenging conditions to treat because of their complexity, severity, and expense. In contrast, it is not essential for the procedure to be successful, as the outcomes of the operation may fail. Alzheimer's disease, which primarily affects older persons, is one of the most prevalent forms of dementia and is associated with memory loss and forgetfulness. Depending on the individual's health status. Because of this, brain CT scans can be used to evaluate a patient's level of Alzheimer's disease and classify his or her memory loss. Neuroimaging data, including MRI scans, have been studied extensively in recent years as a potential diagnostic tool for Alzheimer's disease. Recent years have seen significant computer-based research advancements in the field of DL. Recent days have seen significant progress in the application of deep learning algorithms to the study of medical imaging. We propose a deep convolutional network and show its effectiveness on the Alzheimer's Disease (AD) Dataset we downloaded from Kaggle to accomplish this goal. The best parameters for Alzheimer's disease prediction have been calculated using ResNet 50 and the Xception model. This study aims to categorise AD photos into four categories recognised by neurologists, and the findings will be evaluated using several criteria. In this study, computer methods, namely DCNN and transfer learning, were utilised to categorise AD. ResNet 50 performed best in two of the seven criteria used for the evaluation: accuracy and AUC score. The results of this study conclusively demonstrate that AD may be classified by computer algorithms into four categories recognised by medical professionals. The proposed method produces superior results, with a best-in-class training accuracy of 94.51% and validation accuracy of 86.66% for AD. In comparison to similar works, this the score of accuracy is substantially higher.**

Keywords— ***Alzheimer's disease, Deep learning, ResNet 50, Early stage detection and diagnosis,***

I. INTRODUCTION

The brain is an extraordinarily important and complex organ. Functions as varied as ideation, problem-solving, thinking, decision-making, creativity, and remembering all rely on it. The knowledge or experiences stored in memory can be retrieved later. The experiences and information contained in our bodies' memories are crucial to the development of our personalities and identities. It is terrible to lose one's memory due to dementia and become unfamiliar with one's surroundings. Alzheimer's disease dementia is more common than any other form of the illness[1][2]. People's concerns about getting Alzheimer's grow as age. The steady death of brain cells in Alzheimer's disease causes patients to withdraw emotionally and socially, forget their early years, have trouble recognising familiar faces, and have difficulty following even the most basic directions. In the latter stages, unable to swallow, cough, or breathe. Health and social care for the globe's 50 million dementia sufferers costs as much as the 18th largest economy in the world [3]. Memory loss is the major symptom of the Mild Cognitive Impairment (MCI) stage of Alzheimer's disease, the most common neurodegenerative condition, which is followed by the onset of behavioural problems and a decline in the ability to care for oneself. Although MCI is a warning sign of Alzheimer's disease, not all people who have it go on to get the condition. Some people with MCI stay in the MCI stage indefinitely, whereas others acquire non-AD dementia. Although there is currently no treatment for AD, it is nevertheless important to forecast that patients with MCI may eventually acquire AD.[4][5].

Concurrently, it would be desirable to accurately identify individuals in the MCI stage who do not progress to AD to save them from unnecessary pharmacologic drugs that may be of little assistance & may do more harm through side effects. [5].

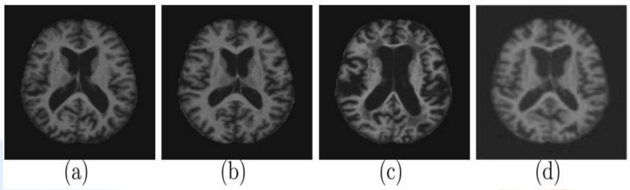

Fig. 1. Various MRI scans of the brain, each depicting a distinct stage of Alzheimer's disease. (a) Nondemented; (b) minor cognitive impairment; (c) memory loss; (d) Dementia, Moderate.

Alzheimer's disease dementia is often classified as one of five stages [6]:

- **Early or Preclinical stage:** There may be no symptoms at this time, or the individual may just be having mild memory problems.
- **Mild Cognitive Impairment (MCI):** Memory loss may be more noticeable at this stage, although the individual may still be able to function independently.
- **Mild Dementia:** At this point, the person may have difficulty recalling recent events, doing routine chores, or expressing themselves clearly. They might also have mood swings, bewilderment, and disorientation.
- **Moderate Dementia:** The person may have trouble doing routine tasks and may have trouble recalling familiar faces at this point. Their memory loss, perplexity, and character shifts may be more extreme.
- **Severe Dementia:** Individual may have lost the ability to communicate with and recognise loved ones, and is completely dependant on others for care at this stage.

Researchers are looking for ways to detect this condition early on in the hopes of slowing the disease's abnormal degeneration of the brain, reducing healthcare expenses, and improving treatment options.[7][8]. Recent failures in trials of Alzheimer's disease therapy may highlight the need of early intervention and diagnosis. The growing importance of dementia diagnosis in a variety of neuroimaging techniques has led to the development of several new diagnostic criteria. When applied to neuroimaging, deep learning improves the diagnostic accuracy of different types of dementia. To apply deep learning algorithms, specific pre-processing processes are required. Features are extracted, selected, and reduced in dimensionality before being used in a classifier algorithm as part of a deep learning-based categorization process. Such approaches need sophisticated expertise as well as multiple optimisation phases that might be time-consuming[9].

DL is a subfield of ML that helps computers get better at their tasks by teaching them from their prior mistakes and successes. Classifying fresh occurrences and predicting novel patterns requires prior learning (training). The effectiveness of deep learning greatly exceeds that of more conventional statistical methods. To implement machine learning successfully, one must have an intimate understanding of both the problem domain and the constraints of the underlying algorithms. If trials are conducted properly, knowledge is used, and results are rigorously verified, then it has a high chance of success.[10].

This study aims to construct a model that can effectively categorise the AD dataset into various classes of dataset, allowing for the detection AD in it early stages using deep leaning models on structural MRI scans in a shorter amount of time. Following is a brief overview of this project's primary goals:

- To create DL technique to recognise various MRI images, such as mild, moderate, nondemented, and extremely mild demented images.
- This research is to use an MRI dataset available on Kaggle to test a chosen ML and DL method's capability to detect the onset of Alzheimer's disease.
- To balance the dataset with SMOTE and optimize the images by removing noise from the MRI image collection in the data preprocessing phase.
- Several data augmentation techniques are employed to improve picture quality.
- A prospective machine learning and deep learning approach were discovered to suggest an improvement for better performance in identifying AD.
- To improve the detection of AD, we built a mixed AI-based model that combines ML & DL
- The purpose of this study is to use numerous performance measures to assess the efficacy of the presented methods for predicting from MRI images.

II. LITERATURE SURVEY

Deep learning models have been developed using a variety of techniques to analyse MRI scans for signs of Alzheimer's disease. Multiple researchers have developed various methods for identifying Alzheimer's disease throughout the years. The sections that follow provide an overview of the progress made thus far.

Menagadevi et al.[11] developed an AI system for the identification of AD by combining a DL model with traditional classification methods. To begin improving the input MRI pictures, they first put them through a series of preprocessing steps. After the pictures have been preprocessed, segmentation is performed to produce the ROI. After that, researchers use the recommended multiscale pooling residual autoencoder algorithm to extract the features. They experimented with a wide variety of classifiers before settling on KNN & Extreme Learning Machine. The total accuracy for the binary classification test with the KNN classifier was 96.88%, while with the ELM classifier it was 98.99%.

Murugan et al.[12] created a DL technique called DEMNET to detect AD in MRIs. Preprocessing, oversampling, & data splitting were only some of the image processing techniques employed. After partitioning the data, they used the provided deep model to extract features and classify the data. On average, they were 95.23 percent accurate while performing a number of different classifications.

Loddo et al. [7] provided an entirely automated strategy for identifying AD in MRI scans by using ensemble deep learning techniques. AlexNet, ResNet 101, and InceptionResNetV2 were the three pretrained deep models used. The ensemble output was then generated using an average approach. They were the most successful group in both the binary classification test (96.57%) and the multi-classification task (97.7%).

Sharma et al.[13] introduced an AI-based hybrid modality (HTLML) for AD detection in MRI scans. They are responsible for the first phase of MRI image preparation. They then simultaneously fed these processed pictures into two pretrained models like DenseNet201 and DenseNet121. After that, they use individual classifiers for each pretrained model to carry out classification. At last, they use a voting mechanism to average the results from all of the classifiers. In a multi-classification test, they had a 91.75 percent success rate.

Hazarika et al. [14] Employing deep neural networks & magnetic resonance imaging, a method was proposed for the categorization of AD. To determine if an MRI was taken from a person with Alzheimer's disease or not, the scans are first preprocessed, then characteristics are retrieved from the segmented brain images using a combination of 2D and 3D CNNs, and lastly the scans are classed. The authors reported encouraging findings, with a sensitivity of 96%, specificity of 94.67%, and an accuracy of 95.34%. This approach has the potential to greatly enhance Alzheimer's disease early identification and treatment.

Tuvshinjargal and Hwang [15] showed an AD MRI prediction model based on a combination of the VGG-C transform and a CNN. They preprocess the incoming photos by quantizing pixel intensity using Z-score scaling. The VGG pretrained model was then used to make predictions on these photos. Their multi-classification task testing accuracy was 77.46 percent.

Balaji et al. [16] suggested a hybrid deep learning method for interpreting MRI scans for signs of AD. The researchers learn spatial & temporal properties from MRI data using a CNN and LSTM. The scientists claim that their hybridised deep learning algorithm successfully classifies MRI images into AD or normal instances with an accuracy of 98.50 percent.

Table 1 summarises the many approaches used so far and lists the benefits and drawbacks of each. Our innovative deep learning model achieves superior performance in binary and multi-classification tasks, overcoming the constraints of all prior work in this area. In the following paragraphs, they will go into depth about our model and the dataset we utilised.

TABLE I. A REVIEW OF THE LITERATURE ON USING DEEP LEARNING TO IDENTIFY AD.

Authors and Reference	Approaches	Results	Limitations
Menagadevi et al. (2023)[11]	Pooling residual autoencoder + ELM	KNN classifier accuracy of 96.88% and overall accuracy of 98.99%	Complex model
Murugan et al. (2021)[12]	Preprocessing + CNN + RMS	accuracy of 95.23%	Overfitting problem
Loddo et al. (2022)[7]	Pretrained models + Ensemble classifier	multi-classification accuracy of 97.7 % and binary classification accuracy of 96.57 %.	Unreliable results across all sizes of datasets
Sharma et al. (2022)[13]	Pretrained models + SVM	overall accuracy of 91.75% for the multi-classification task.	Low accuracy with big data
Hazarika et al. (2023) [14]	Preprocessing + 2D CNN and 3D CNN	high precision (95.34%), high sensitivity (96%), & high specificity (94.67%)	Results on a binary classification challenge were subpar.
Tuvshinjargal and Hwang [15]	Preprocessing + pretrained models	precision in every single test: 77.46	Lackluster results on unbalanced data sets.

Balaji et al. (2023)[16]	3D CNN + LSTM	MRI images classified with a 98.50% degree of accuracy	Complex characteristics can only be learned correctly with a substantial quantity of data.

However, there are restrictions on how deep learning may be used for AD detection. Factors including dataset quality and size, model design, and parameter optimization may all have an impact on how well a deep learning model performs. Further study is required to completely understand the capabilities and limits of deep learning in the identification of AD, notwithstanding these limitations. The goal of this research is to give a critical analysis of the limitations of existing literature on the use of deep learning to AD detection. Furthermore, we address these shortcomings by suggesting a novel, low-overhead deep learning model for robust AD identification in MRI scans. Our model's reliability and performance were guaranteed by a multi-pronged training strategy that included data augmentation, DL, and early stopping. Accuracy, precision, recall, and the F1-score were only few of the measures they used to assess our model's efficacy.

III. Research Methodology

In this part, we'll go through the dataset that was utilised and provide you some instances of the data in visual form. Furthermore, this part discusses all hyperparameter settings for the whole proposed model.

A. *Problem Statement*

The challenge statement is to identify Alzheimer's disease (AD) by designing and training a deep convolutional neural network capable of reliably diagnosing the condition using the patients' MRI medical data. The patient's data includes a person's name and date of birth to results from neuropsychological and cognitive tests, MRI images and even the DNA of the patients. The model's predictions must be accurate, sensitive, and precise in order to distinguish between healthy groups and Alzheimer's disease patients. Alzheimer's disease progresses through four stages: very mild, mild, moderate, and severe. However, getting an accurate diagnosis and categorization of AD is essential for starting therapy early enough to slow or stop brain tissue degradation. Medication for AD must be taken on a regular basis in order to maintain control. A wide variety of medical imaging modalities have benefited greatly from the adoption of deep learning models for analysis. Notable findings for organ and substructure segmentation, diagnosis of severe serious diseases, and classification have been achieved using deep models in the disciplines of pathology, brain, lung, abdomen, heart, breast, bone, retina, and so on. However, the use of DL models for the diagnosis of Alzheimer's disease has received very little attention. [5]. DCNN is preferable to other techniques for obtaining features due to its high level of feature extraction. DCNN combine and train both feature extraction and classification networks simultaneously. In addition to traditional methods of medical picture categorization and computer-assisted diagnosis, deep learning methods were also included into these tools. However, appropriate pre-processing procedures must be implemented prior to utilizing such methods. In addition, for classification and prediction, these techniques require the extraction of features, selection of features, reduction of dimensionality, and feature-based classification. These processes need for not just specialist expertise but also a lengthy series of optimization phases.

B. *Proposed Methodology*

ML & DL are becoming increasingly important as technology advances and more data is collected by brain-imaging methods for extracting useful information and developing reliable predictions of AD from this data. As a result, ML based solutions were proposed to address the shortcomings of the conventional approach to disease prediction [3]. In order to automatically classify Alzheimer's disease, this study plans to assess DL-based MRI feature extraction.The performance of the model was tested between completely connected layers as it was designed as a DL technique for Alzheimer's disease diagnosis on MRI images using several classifiers. Using Deep learning methods, Alzheimer's disease MRI image categorization is the focus of this study. A dataset namely Alzheimer Disease (4 class of images) was obtained from the official website of Kaggle. The dataset includes four disease stages: non-mild, very mild, mild, and moderate. After that data is pre-processed with different steps with resizing the images into 128*128,convert data into array and data balancing using SMOTE and data labeling . The dataset is then split into three parts: validation data (20%), test data (20%), and training data (60%). Then apply two deep learning model that is ResNet 50 and Xception model with some hyperparameter tuning is used for feature extraction and same are used for producing higher accuracy for better prediction of the disease. After this, the performance of the model is evaluated using a variety of measures, including Recall, F1-Score, Precision, ROC, and Accuracy score. The process of research methodology provided below subsection also shows in figure 2.

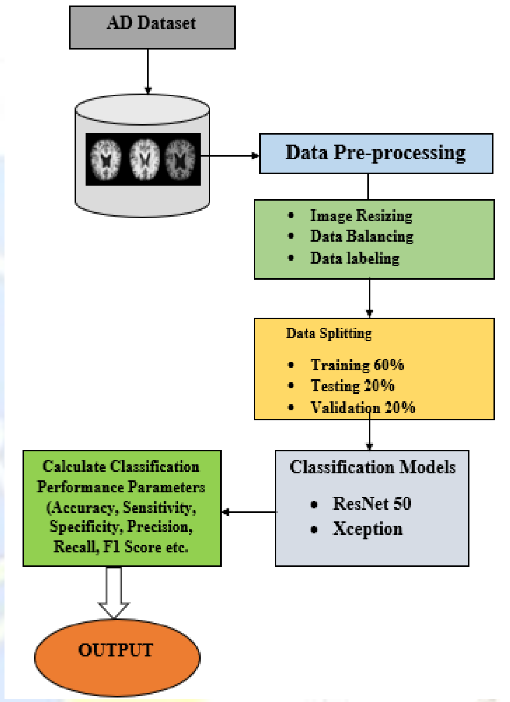

Fig. 2. Flow Chart for Proposed Methodology

a) Data Collection

The Alzheimer's Dataset (4 Image Classes) was utilised for this analysis. There are a total of 6,400 photos in the dataset, which was split up as follows: There are a total of 4098 photos and 4 classes in the train set, 1279 images and 4 classes in the test set, and 1023 images and 4 classes in the validation set. Mild dementia, moderate dementia, no dementia, and very mild dementia are the four phases of Alzheimer's disease.

b) Data preprocessing

Second, the gathered MRI dataset undergoes pre-processing. Knowledge discovery relies heavily on high-quality input data, making data preparation a crucial first step. Huge dividends for decision making may be expected from the detection of data anomalies, the correction of mistakes, and the reduction of data to be analyzed. In this study the main objective of the data pre-processing is process the input image data. To minimize the amount of data, discover relations within the data, normalize the data so that outliers are removed, and extract features. In this particular research project, some steps of performed such as data resizing, convert data into array, data labeling and data balancing using SMOTE.

- **Resizing image:** In computer vision, resizing images is an important first step in the processing pipeline. In general, deep learning models could be trained more quickly with smaller images. The training period is lengthened since the bigger input picture has four times as many pixels as the smaller one.[17]. The picture is downscaled to 128x128 pixels at this point in the data preprocessing phase.
- **Data labeling :** In order for a deep learning model to make sense of raw data (pictures, text files, videos, etc.), the data must be recognised and labelled with one or more relevant and meaningful labels. This research categorises the four phases of Alzheimer's disease progression into (0,1,2,3). The labelled data categories are shown in table 2.

TABLE II. FOUR DEVELOPMENT STAGES OF AD DISEASE

Data Categories	Labels
Non-Demented	0
Very Mild Demented	1
Mild-Demented	2
Moderate-Demented	3

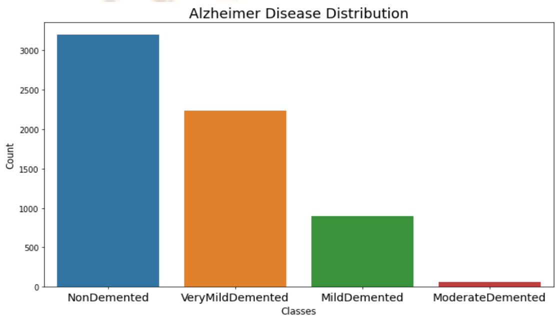

Fig. 3. Data distribution graph before using smote

Figure 3 depicts data distribution graph before smote of all Alzheimer's disease classes. The graph's Classes, which range from "not demented" through "mildly demented" to "moderately deranged," are labelled along the x axis. The y axis of the graph ranges from 0 to 300 to depict the total number of pictures. More pictures may be found in the "Non demented" category.

c) Data balancing using SMOTE

Researchers and practitioners doing classification tasks on unbalanced datasets have difficulties due to the propensity of categorising imbalanced datasets to aggressively designate minority labels as majority class. Because of this, the results of the classification job suffer from increased false positives. Given the scarcity of data for the minority group, researchers and practitioners have the difficulty of avoiding too confident predictions for the majority group.

This has led to the development of a number of oversampling methods, which have become standard practise prior to implementing any classification assignment when unbalanced datasets are present. The SMOTE method is a popular and well-known oversampling strategy. Another resampling method used for balancing datasets with a severely uneven ratio is the Synthetic Minority Oversampling Technique (SMOTE), which tries to enhance the proportion of minority class samples by producing synthetic samples in the minority class. The synthetic production of fresh samples deviated from the multiplication procedure to prevent overfitting.[18].

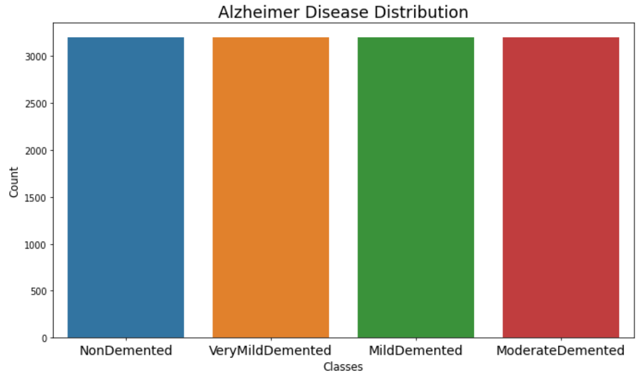

Fig. 4. Data distribution graph after using smote

The SMOTE method's effect on the data-balancing graph is seen in Figure 4. Class descriptors, such as "non-demented," "very mild demented," "mildly demented," and "moderately demented," may be seen along the x axis, while the total number of photos belonging to each class can be seen along the y axis, which ranges from 0 to 300. When using smite, there is parity between all classes.

d) Data Splitting

To ensure the final model is accurate, it is compared to both the training and test data. Machine learning often requires data to be partitioned into at least three distinct groups. The images used in this research are split into three categories: training (60 %), testing (20 %), & validation (20 %).

e) Classification Models

Deep learning requires large amount of data to train the machine so that it can generate accurate results so training time of DNN is more than traditional machine learning models[19]. In this study Resnet 50 and xception model is used.

- **ResNet 50 Model**

CDR was classified using ResNet-50 with data from MRI images only. Keras (TensorFlow backend) was used for this model. ResNet-50 uses a residual deep learning network to solve the problem of convolutional neural network (CNN) gradients vanishing during back-propagation (with 50 layers). He, Zhang, and Sun developed the ResNet-50 model, and an ensemble of ResNet models with varied depths has been victorious in image classification contests. As long as over-fitting is avoided, a deeper network should provide better results. The signal required to adjust the weights decreases dramatically at the early layers as the network's depth grows; this is because the final layers of the network compare ground-truth and prediction. This means that the information presented in prior levels is effectively unlearned. When attempting to optimise weights via nonlinear optimisation, the gradient (a matrix of second order derivatives) approaches zero, leading to the dreaded "vanishing gradient" issue. The second problem with training deeper networks is that merely adding more layers will raise the training error without any further thought. Residual networks, which construct the network using modules called residual models, make it feasible to train such deep networks. The term for this issue is "degradation." The ResNet-50 model relies heavily on convolutional blocks. In order to identify pictures, these networks employ several filters (for instance, a filter with a size of 3 by 3 pixels). These filters are walked over the base picture in steps. The picture values are multiplied by the values in the filters (which are also learnt). Downsampling is achieved while the most salient characteristics are retained by pooling the outputs of these filters (e.g., the maximum values recovered after applying filters) [20].

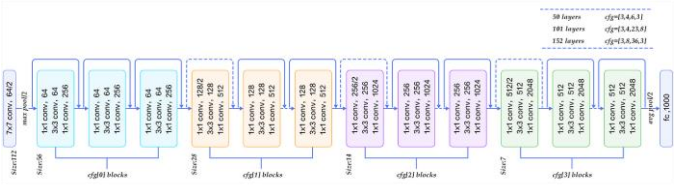

Fig. 5. General Architecture of Res net 50 model

- **Xception Model**

The Xception model is a neural network model based on deep learning that was inspired by Inception. using an adjusted convolution layer that can distinguish between different depths. The Xception architecture uses 36 convolutional layers for feature extraction. Each of these layers is organised into 14 modules, with the exception of the first and last units, which form the network's feature extraction rule. In 2017, François Chollet presented the Xception deep learning model architecture. For the purpose of classifying images, a CNN architecture was used. Although the Xception model uses a more severe kind of depth wise separated convolutions than the Inception design, it is still based on that architecture. The key innovation of Xception, which stands for "Extreme Inception," is how convolutions are handled. By dividing the procedure into two steps—depth wise the convolutions and pointwise convolutions—Xception avoids the drawbacks of utilizing traditional convolutions. Pointwise convolutions aggregate the results of depth wise convolutions by performing 1x1 convolutions, while depth wise the convolutions deploy a single convolutional filter to each input channel. Comparing Xception's convolutional layers to this separation of the convolutions provides for a reduction in computational cost. The precision is maintained or even increased while fewer parameters and processes are required to accomplish this. Object identification, segmentation, and picture classification are just a few of the computer vision tasks for which Xception has been extensively employed. The Xception model may be trained on a labelled dataset of MRI images to categorise them into distinct categories or detect certain disorders or anomalies when it comes to MRI image classification[21].

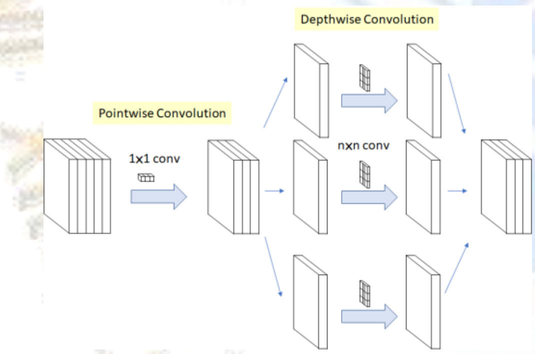

Fig. 6. General Architecture of Xception model

C. Proposed Algorithm

The proposed algorithm for carrying out the study is shown below:

Input: Alzheimer's Disease MRI dataset

Output: Prediction

1) Collecting and loading the Alzheimer's Disease dataset obtained from Kaggle.

2) Import Python libraries such as numpy, pandas, tensorflow, keras, open-cv, matplotlib etc
3) Pre-processing of data.
   - Resizing 128*128
   - Data balancing
   - Convert data into array
   - Data labeling
4) Data will be divided as follows: 60% training data, 20% test data, and 20% validation data.
5) Applying SMOTE for data balancing
   - Epochs = 20
   - Batch_size = 64
   - Loss function = Categorical_crossentropy
   - Optimizer = Adam(learning_rate=1e-3)
6) Creating ResNet 50 and Xception Model

7) Accuracy, sensitivity, specificity, F1-score, precision, area under the curve (AUC), and recall may be used to evaluate a model's efficacy.
8) Result

IV. Result Analysis and Discussion

To train the aforementioned AD deep model, many experiments are conducted utilising the Kaggle dataset and the suggested deep model. The suggested AD detection algorithm was evaluated using a standard training and testing procedure to provide a fair and accurate assessment. Both the ResNet 50 and Xception methods rely on a GPU (in this case, an NVIDIA Tesla T4 GPU) and 14 GB of DDR4 RAM to run well. The solution makes use of the Python library, which includes such packages as Keras, pandas, nampy, seaborn, matplotlib, etc. Both models were trained using the 'Adam' optimizer, although the loss functions for Model 1 and Model 2 are of different types. Model 1's loss function is binary cross-entropy, whereas Model 2's is Categorical. The results of this investigation were evaluated using four different metrics: accuracy, precision, recall, and F1-score.

A. Evaluation Parameters

Performance metrics are critical in the creation, selection, and assessment of machine learning models. There are several classification measures, including Accuracy, Recall, F measure, and Area Under the ROC Curve [22]. Standard criteria obtained from the generated confusion matrix were used to compare the efficacy of different ML systems for predicting AD.

Confusion Matrix: In order to measure the efficiency of ML algorithms, a special table known as the confusion matrix is used. A data matrix, in which each row represents a real-world class and each column an expected-world class or vice versa, can be thought of as a sample library from which to draw. The confusion matrix keeps track of every possible result of a test, including the number of false positives, false negatives, true negatives, and true positives.

- $\boldsymbol{Tp}$ = number of confirmed positive cases
- $\boldsymbol{Fp}$ = number of false-negative cases
- $\boldsymbol{FN}$ = number of false-positive cases
- $\boldsymbol{TN}$ = number of true negative cases

It was concluded that the most crucial elements to take into account while assessing the efficacy of our model were precision, accuracy, ROC, recall, and F1 score:

Accuracy: Accuracy (ACC) is the proportion of examined instances with correct answers (TP and TN). The most precise value is 1, while the least reliable value is 0. To calculate ACC, one can use the following (1) approach.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \dots . (1)$$

Recall: The recall rate may be calculated by dividing the number of accurate predictions by the total number of forecasts. One recall is the ideal number, while zero recalls is the worst. To determine recall, people often utilise the following approach:

$$R = \frac{TP}{TP + TN} \dots . (2)$$

Precision: The ratio of correct guesses to the total number of correct predictions is the definition of accuracy. Precision may be calculated as follows, with 1 representing the most accuracy and 0 representing the lowest.

$$Precision = \frac{TP}{TP + FP} \ldots . (3)$$

F1 score: Recall and accuracy are weighted harmonic averaged to get the F-measure. The Recall and Precision measurements are frequently combined into one measurement when evaluating different machine learning algorithms. Presenting the F-measure formula are:

$$F1 - score = 2 \times \frac{precision * recall}{precision + recall} \ldots . (4)$$

AUC and Receiver Operating Characteristics (ROC): ROC graphs show a receiver's effectiveness. The true positive ratio (TPR) and false positive rate (FPR) are contrasted along the x-axis in the 2D graph known as the ROC curve. In order to demonstrate how classifiers, distinguish between categories, lines are drawn between the thresholds that are chosen when choosing in binary classification.

With values ranging from zero to one, the area under the curve (AUC) is a common statistic used with ROC curves. An AUC greater than 0.5 implies that a classifier is well-trained when it is giving more weight to correct forecasts and less to incorrect ones. A badly trained classifier will have an AUC close to 0.5 and a ROC curve that is a diagonal line. Area Under the ROC Curve, or AUC for short. The area under the whole ROC curve is what AUC measures.

B. *Experimental Results*

Major findings after using the proposed technique have been presented in this section. It shows the all obtained results for the AD images dataset. The classification is performed into two models are given below:

a) *Results of proposed ResNet-50 Model*

Results from simulations using the proposed ResNet-50 Model are provided in this section.

TABLE III. TRAINING ACCURACY AND LOSS RESULTS OF RESNET 50 MODEL

Model	Training Accuracy	Train Loss	Val Accuracy	Val Loss
ResNet 50	94.51	27.68	87.66	67.76

The above table 3 shows the Training and Validation results of ResNet 50 model. Where the results show that the model has training accuracy of 94.51 with loss of 27.68. Whereas, the model has validation accuracy of 87.66 and validation loss of 67.76.

TABLE IV. MODEL PERFORMANCE EVALUATION RESULTS OF RESNET 50

Particular	Precision	Recall	AUC	F1_Score
Training	90.10	87.68	98.64	88.87
Testing	76.17	73.68	93.28	74.86

The above table 4 are the training and testing results of ResNet 50 model in terms of precision, recall,AUC score and F1 score. Where the training results of model shows precision is 90.10, recall 87.68, AUC 98.64 and F1 Score 88.87. On the other hand, Testing results shows the precision of 76.17, recall 73.68, AUC 93.28 and F1 Score 74.86. From the above results of training and testing it can be concluded that the model has performed well in training the model with precision, recall, auc and f1 score.

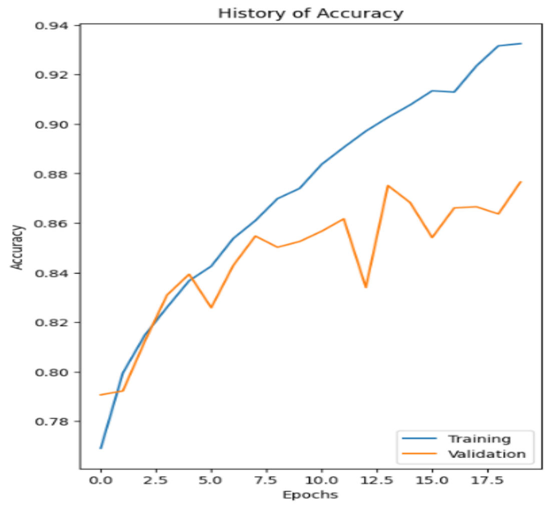

Fig. 7. Accuracy Graph of Resnet 50 Model

The accuracy graph of the suggested Resnet 50 model is shown in Figure 7. The precision ranges from 0.78 to 0.94 along the y-axis, while the number of epochs ranges from 0 to 17.5 along the x-axis. Training accuracy is shown by the blue line, whereas validation accuracy is shown by the orange line. The suggested model achieves 94.51% accuracy during training and 86.66% during validation.

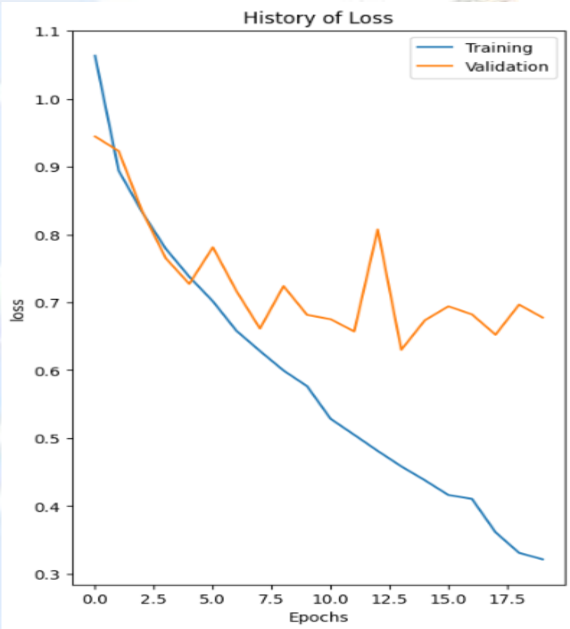

Fig. 8. Loss graph of Resnet 50Model

The loss graph for the Resnet 50 model is shown in Figure 8. Along the x-axis is the number of epochs, and along the y-axis are the corresponding loss values. Validation loss is shown in orange, whereas training loss is shown in blue. In the loss graph, training yields a loss of 0.2768, whereas validation yields a loss of 0.6776.

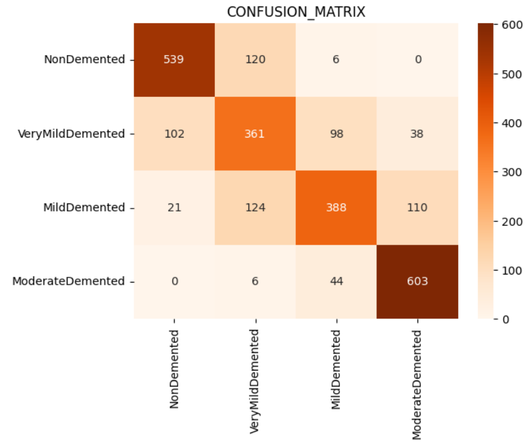

Fig. 9. Confusion Matrix of ResNet 50 Model

The Resnet 50 model's labelled confusion matrix is shown in Figure 9. Labels such as "non-demented," "very light demented," "mildly demented," and "moderately demented" are shown in this confusion matrix. This matrix depicts the many types of categorization. The correctly predicted values are given in this diagonal and other values are incorrectly predicted

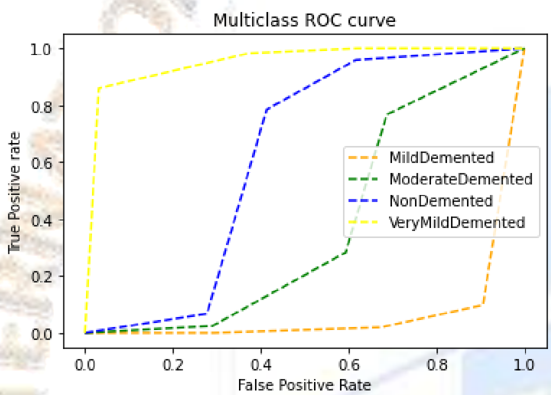

Fig. 10. Multiclass ROC Curve of Resnet 50 model

The area under the receiver operating characteristic curve is shown by the dashed line in Figure 10. The x-axis represents the proportion of false positives, the y-axis the proportion of true positives, and the yellow dotted line the prevalence of very mild dementia, the blue line the absence of dementia, the green line the prevalence of moderate dementia, and the orange line the prevalence of mild dementia.

b) Results of proposed Xception Model

In this section provide the simulation results of proposed Xception Model.

TABLE V. TRAINING ACCURACY AND LOSS RESULTS OF XCEPTION MODEL

Model	Training Accuracy	Train Loss	Val Accuracy	Val Loss
Xception	89.22	0.5053	83.75	0.7835

The above table 5 shows the Training and Validation results of Xception model. Where the results show that the model has training accuracy of 89.22 with loss of 0.5053. Whereas, the model has validation accuracy of 83.75 and validation loss of 0.7835.

TABLE VI. MODEL PERFORMANCE EVALUATION RESULTS OF XCEPTION 50

Particular	Precision	Recall	AUC	F1_Score
Training	80.55	74.96%	95.32	77.61
Testing	69.06	63.43	89.79	66.05

The above table 6 are the training and testing results of Xception model in terms of accuracy, precision,recall and f1 score where the training results of model shows precision 80.55%, recall 74.96%, AUC 95.32% and F1 Score 77.61%. On the other hand, Testing results shows the precision of 69.06%, recall 63.43, AUC 89.79% and F1 Score 66.05%. From the above results of training and testing it can be concluded that the model has performed well.

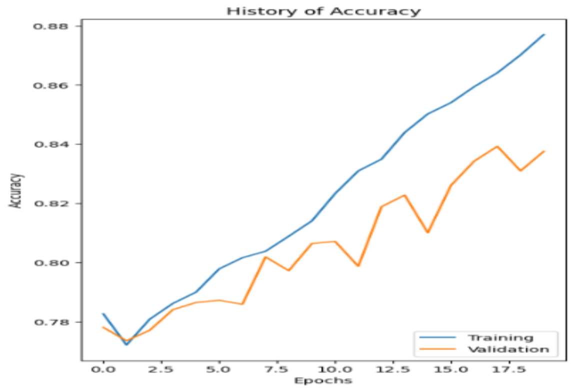

Fig. 11. Accuracy graph of Xception model

The suggested Xception model accuracy graph is shown in Figure 11. The accuracy values range from 0.78 to 0.88 along the y-axis of this graph, while the x-axis shows the number of epochs, from 0 to 17.5. The training accuracy is represented by the blue line, and the validation accuracy by the orange line. The suggested model achieves an accuracy of 89.22% in training and 83.75% in validation.

Fig. 12. Loss graph of Xception model

Figure 12 displays the loss graph for the Xception model. The epoch count is shown along the x-axis, while the loss rate is indicated along the y-axis. The training loss is shown in blue, while the validation loss is shown in orange. When plotting the loss, validation has a loss of 0.7835 whereas training has a loss of 0.5053.

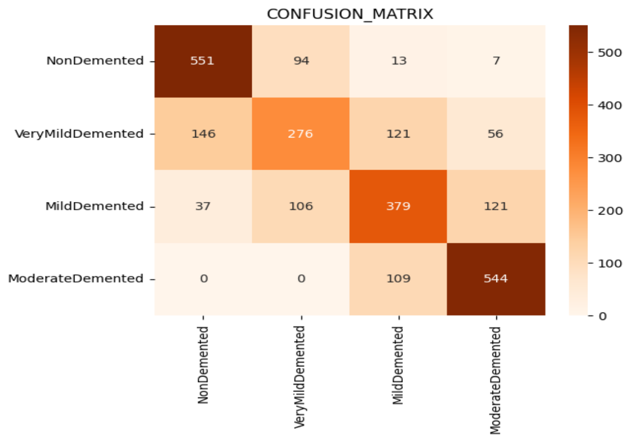

Fig. 13. Confusion Matrix of xception model

Figure 13 displays the labelled confusion matrix from the xception model. This confusion matrix represents a number of different types. This matrix demonstrates the multiple classifications. This diagonal contains the accurately predicted values, while other values are incorrectly predicted.

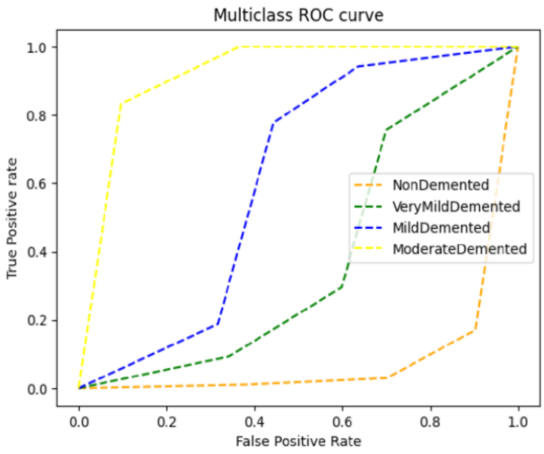

Fig. 14. Figure 14: Multiclass ROC Curve of xception model

The above Fig. 14 represents the Xception Model Multiclass ROC curve. The number of dataset classes are shown in the curve, including them yellow line shows the moderate demented, blue line shows the mild demented, green line shows the very mild demented and orange lines shows the non-demented. False positive rates are shown on the x axis while genuine positive rates, which span from 0 to 1, are plotted on the y axis.

c) Comparison between base and propose Models

In this section provide the simulation results of proposed Xception and ResNet50 or base Vgg-19 Models.

TABLE VII. PERFORMANCE COMPARISON BETWEEN BASE AND PROPOSED MODEL

Models	Train accuracy	Train loss	Validation accuracy	Validation loss
Vgg19(base)	0.7156	27.1876	0.7058	26.749
Resnet 50	0.9451	0.2768	0.8766	0.6776
Xception model	0.8922	0.5053	0.8375	0.7835

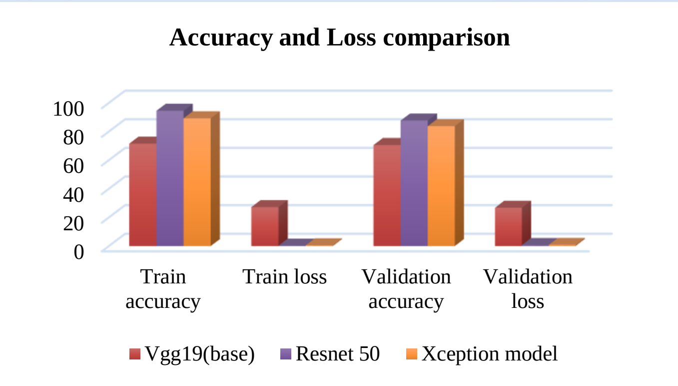

Fig. 15. Accuracy Comparison Between Base and Proposed Model

Accuracy and loss during training and testing for the proposed and the base model are shown in Figure 15 and Table 7. There was a 94.51% training accuracy and an 876.6% validation accuracy for the suggested model (Resnet 50), whereas the corresponding figures for the xception model were 89.22% and 83.75%, respectively. When compared to the suggested model, the base model's (Vgg190) training accuracy is only 71.56 percent, and its validation accuracy is just 70.58 percent. As a result, the suggested model has improved upon the accuracy of the previous model. Training loss for the proposed Resnet 50 is 0.2768, whereas validation loss is 0.6776 for the xception model, and training loss for the base model is 27.1876 and validation loss for the base model is 26.7495. It can say that the loss of proposed model is very lower than the existing model.

TABLE VIII. COMPARISONS PERFORMANCE RESULT OF TESTING BETWEEN BASE AND PROPOSED MODELS

Models	Precision	Recall	AUC Score	F1_Score
Vgg19(base)	0.4115	0.4111	0.6168	0.4114
Resnet 50	0.7617	0.7368	0.9328	0.7486
Xception model	0.6906	0.6343	0.8979	0.6605

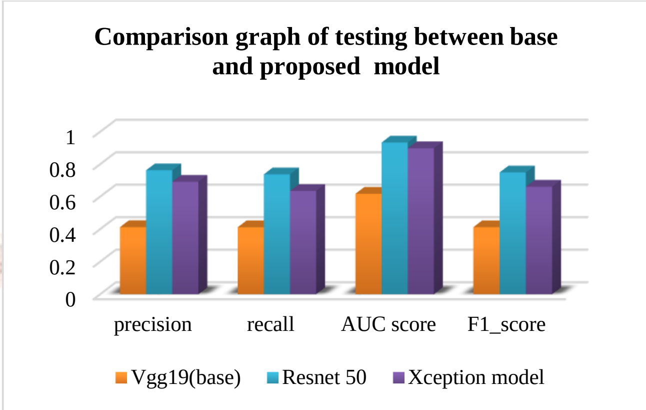

Fig. 16. Comparisons result of testing between base and proposed models

Figure 16 and Table 8depict the testing results comparing the base model to the proposed model in terms of accuracy, recall, f1 score, and area under the curve. The proposed model (Resnet20) achieves 76.17% precision, 73.68% recall, 93.28% Auc score, and 74.86% F1 score value, while the xception model achieves 69.06% precision, 63.43% recall, 89.79% Auc score, and 66.05% F1 score value, and the base model achieves 41.15%, precision, 41.11% recall, 61.68%, Auc score, and 41.14% f1 score value.

V. CONCLUSION AND FUTURE WORK

Dementia is most often caused by Alzheimer's disease. Alzheimer's disease is a progressive, fatal brain illness for which there is now no effective treatment. However, current treatment options may slow its progression. Therefore, stopping and slowing the development of AD requires its early identification. The primary goal is to develop an end-to-end framework for medical image categorization across all phases of Alzheimer's disease. In this study, we use a deep learning technique known as CNN. In this work, we evaluate a possible approach to early diagnosis. Overall, the models used in this work have performed well, correctly categorising the photos into four distinct groups. We find that the performance of ResNet-50 and Xception exceeds that of the state-of-the-art models. The suggested model's efficacy has been measured in terms of its recall, sensitivity, and precision. According to the results, shows proposed model (Resnet 50) achieved training accuracy of 94.51 % and xception model achieve the training accuracy 89.22 % and proposed model (Resnet20) precision is 90.10%, recall is 87.68%,Auc score 98.64% and F1 score value is 88..87% and xception model show a precision value is 80.55%,recall is 74.96%,Auc score 95.32% and F1 score value is 77.61%.This shows that the proposed model is powerful, effective, energetic, and skilled for the Classifying and detect AD. In order to utilise this model in clinical settings and increase the health care rate against this illness, further study is needed. People should be made aware of the existence of this illness and urged to undergo screenings. For more direct application, we are working on putting this model up on a website. A bigger dataset will be available for future testing of this model. For the 'Moderate Demented' class, the present dataset only provided a small number of photos for use in training and evaluation. The suggested algorithm may aid in the accurate diagnosis of Alzheimer's disease and might be expanded to automatically detect additional neurodegenerative illnesses in the future. Therefore, in the future, neuroimaging methods may include quicker, more accurate, and more efficient classification algorithms, allowing the generation of a diagnostic hypothesis from a single brain scan.

REFERENCES

[1] M. A. Deture and D. W. Dickson, "The neuropathological diagnosis of Alzheimer's disease," *Molecular Neurodegeneration*. 2019. doi: 10.1186/s13024-019-0333-5.

[2] H. Kocaelli, M. Yaltirik, L. I. Yargic, and H. Özbas, "Alzheimer's disease and dental management," *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.*, 2002, doi: 10.1067/moe.2002.123538.

[3] 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, 2022, doi: 10.3390/s22082911.

[4] T. J. Saleem *et al.*, "Deep Learning-Based Diagnosis of Alzheimer's Disease," *J. Pers. Med.*, vol. 12, no. 5, 2022, doi: 10.3390/jpm12050815.

[5] J. Islam and Y. Zhang, "A Novel Deep Learning Based Multi-class Classification Method for Alzheimer's Disease Detection Using Brain MRI Data," 2017. doi: 10.1007/978-3-319-70772-3_20.

[6] R. Kalaria, "Similarities between Alzheimer's disease and vascular dementia," 2002. doi: 10.1016/S0022-510X(02)00256-3.

[7] A. Loddo, S. Buttau, and C. Di Ruberto, "Deep learning based pipelines for Alzheimer's disease diagnosis: A comparative study and a novel deep-ensemble method," *Comput. Biol. Med.*, 2022, doi: 10.1016/j.compbiomed.2021.105032.

[8] P. R. Solomon and C. A. Murphy, "Early diagnosis and treatment of Alzheimer's disease," *Expert Review of Neurotherapeutics*. 2008. doi: 10.1586/14737175.8.5.769.

[9] Z. Zhang, "Improved Adam Optimizer for Deep Neural Networks," 2019. doi: 10.1109/IWQoS.2018.8624183.

[10] V. Patil, M. Madgi, and A. Kiran, "Early prediction of Alzheimer's disease using convolutional neural network: a review," *Egyptian Journal of Neurology, Psychiatry and Neurosurgery*. 2022. doi: 10.1186/s41983-022-00571-w.

[11] M. Menagadevi, S. Mangai, N. Madian, and D. Thiyagarajan, "Automated prediction system for Alzheimer detection based on deep residual autoencoder and support vector machine," *Optik (Stuttg).*, 2023, doi: 10.1016/j.ijleo.2022.170212.

[12] S. Murugan *et al.*, "DEMNET: A Deep Learning Model for Early Diagnosis of Alzheimer Diseases and Dementia from MR Images," *IEEE Access*, 2021, doi: 10.1109/ACCESS.2021.3090474.

[13] S. Sharma *et al.*, "HTLML: Hybrid AI Based Model for Detection of Alzheimer's Disease," *Diagnostics*, 2022, doi: 10.3390/diagnostics12081833.

[14] R. A. Hazarika *et al.*, "An Approach for Classification of Alzheimer's Disease Using Deep Neural Network and Brain Magnetic Resonance Imaging (MRI)," *Electron.*, 2023, doi: 10.3390/electronics12030676.

[15] B. Tuvshinjargal and H. Hwang, "VGG-C Transform Model with Batch Normalization to Predict Alzheimer's Disease through MRI Dataset," *Electron.*, 2022, doi: 10.3390/electronics11162601.

[16] P. Balaji, M. A. Chaurasia, S. M. Bilfaqih, A. Muniasamy, and L. E. G. Alsid, "Hybridized Deep Learning Approach for Detecting Alzheimer's Disease," *Biomedicines*, 2023, doi: 10.3390/biomedicines11010149.

[17] S. Saponara and A. Elhanashi, "Impact of Image Resizing on Deep Learning Detectors for Training Time and Model Performance," 2022. doi: 10.1007/978-3-030-95498-7_2.

[18] A. Fernández, S. García, F. Herrera, and N. V. Chawla, "SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary," *Journal of Artificial Intelligence Research*. 2018. doi: 10.1613/jair.1.11192.

[19] R. Kumari and G. Jambheshwar, "A Review of Deep Learning Techniques," vol. 9, no. 8, pp. 2320–2882, 2021.

[20] L. V. Fulton, D. Dolezel, J. Harrop, Y. Yan, and C. P. Fulton, "Classification of alzheimer's disease with and without imagery using gradient boosted machines and resnet-50," *Brain Sci.*, vol. 9, no. 9, 2019, doi: 10.3390/brainsci9090212.

[21] H. R. Almadhoun and S. S. Abu-naser, "Classification of Alzheimer ' s Disease Using Traditional Classifiers with Pre-Trained CNN," vol. 5, no. 4, pp. 17–21, 2021.

[22] M. Gong, "A Novel Performance Measure for Machine Learning Classification," *Int. J. Manag. Inf. Technol.*, 2021, doi: 10.5121/ijmit.2021.13101.