1 Anmar Mahmood

2 Mesut ÇEVİK

Examining the Potential of Deep Learning in the Early Diagnosis of Alzheimer's Disease using Brain MRI Images

Journal of Electrical Systems

Abstract: - Alzheimer's disease is a severe public health problem affecting millions worldwide. Deep Learning (DL) models can aid in detecting the disease using MRI data, and we evaluated three DL models for this purpose. We used detailed MRI images of Alzheimer's patients and healthy controls to train these models. A convolutional neural network (CNN) with two convolutional and two fully connected layers was employed in the initial model, which had a 95% accuracy rate. The second model, which included a leaky ReLU activation function, more fully connected layers, and a bigger kernel size, was an enhanced version of the previous one and had a 96% accuracy rate. The third model was a transfer learning model with two dense layers built on top of the VGG16 architecture, achieving an accuracy of 80%. Our findings imply how neural network models may assist with MRI data-based the disease assessment via evaluations of reliability, precision, recollection, and the F1 ranking. For enhancing the precision and usability of these gadgets for therapeutic usage, more study must be conducted.

Keywords: Deep Learning, Convolutional Neural Network, Transfer Learning, Alzheimer's illness

I. INTRODUCTION

A deteriorating neurological ailment called Alzheimer's Disease (AD) disrupts recall, thinking, and behavior. This is no treatment, notwithstanding knowing that it's the least typical root of senile in seniors. The progression of Alzheimer's begins with minor memory loss and difficulties in problem-solving, but it gradually worsens, leading to difficulties in performing everyday tasks and communication impairments [1].

Although the exact etiology of AD is still unclear, a mix of inherited, environmental, and lifestyle factors are thought to be responsible [2].

Brain imaging studies have shown that the brain of an Alzheimer's patient is smaller than that of a healthy person, with smaller regions in memory, language, and behavior-related areas. Over time, the brain shrinks, and cells die, leading to further damage [2].

While three medications are approved for treating AD symptoms - cholinesterase inhibitors, NMDA receptor antagonists, and a combination - they do not cure Alzheimer's and become less effective over time. However, they can help slow down the progression of the disease, improve memory and thinking, and increase the overall quality of life [3].

Lifestyle changes can also help manage AD symptoms and improve overall health, including staying mentally and physically active, eating a balanced diet, managing stress, and getting adequate sleep. Caregivers also play a crucial role in supporting those with AD by helping with daily tasks, providing emotional support, and managing symptoms. Support groups, respite care, and community resources can also be helpful for caregivers [4]. There is still much work to be done to find a cure; new treatments and methods for managing its symptoms are being investigated. In the meantime, providing support, care, and understanding to those affected by Alzheimer's and their families is crucial [4].

Dementia monitoring in the cerebral cortex Magnetic resonance imaging (MRI)

Dementia monitoring in the cerebral cortex Magnetic resonance imaging (MRI

With the goal to effectively manage dementia (AD), prompt detection via magnetic resonance imaging (MRI) has grown more essential. MRI is a powerful tool for visualizing brain structure and detecting subtle changes that may indicate the presence of AD. Reduction in the size of the hippocampus, a critical brain region for memory formation, is a key imaging marker of AD and can be detected on MRI scans. Advanced MRI techniques, such as structural MRI and diffusion tensor imaging (DTI), can detect additional changes in brain structure that may indicate AD [5].

One significant advantage of MRI for AD detection is that it is non-invasive and well-tolerated. Additionally, MRI is highly accurate and can detect subtle changes in the brain that may not be apparent on clinical assessments, allowing for earlier detection and intervention. Machine learning algorithms trained on large datasets of MRI scans can identify patterns and relationships between brain imaging markers and AD diagnosis, allowing for highly sensitive and specific detection of AD [6].

However, MRI for early AD detection has limitations and challenges. The cost of MRI scans can be prohibitive, and not all individuals can access this technology. Additionally, the interpretation of MRI scans requires

[1] dep. Electricity and Computer Engineering, Altınbaş University, Diyala, Iraq. anmar.sh.mahmood@gmail.com
[2] dep. Electrical and Electronics Engineering, Altınbaş University, Istanbul, Turkey. mesut.cevik@altinbas.edu.tr

specialized expertise. Finally, while MRI is a sensitive marker of AD, it is not always specific, and other conditions can also cause changes in brain structure.

Despite these challenges, the application of MRI for early detection of AD holds substantial promise for enhancing the prognosis and running of the disease. Further research have being necessitate to fully realize the potential of MRI for early AD detection and ensure that this technology is accessible and affordable for all individuals who may benefit from it [6].

Motivation: The motivation for this work is the urgent need for early dementia diagnosing and identifying diseases, a severe public health problem affecting millions of individuals worldwide. Early analysis is critical for the compelling treatment and the executives of the sickness, and Profound Learning (DL) models have shown potential for helping with identifying Alzheimer's utilizing X-ray information. This study evaluated three DL models for this purpose and achieved promising results with high precise, exactitude, anamnesis, and F1-outcome measurements. The work adds to creating compelling and open instruments as long as the prompt recognition and analysis of Alzheimer's sickness, which can at last prompt improved results for patients and their families.

Here describes how the article is set up:

- Introduction: Provides historical background on dementia, emphasizes the importance for prompt detection, draws an end result, then describes the use of Dramatic Training (DL) algorithms to enhance the diagnosis of dementia using radiation evidence. In addition, this articulates the research objectives.
- Methodology: Depicts the dataset used for getting ready and testing the DL models and the nuances of the model's plan and getting ready interaction.
- Results: Presents the performance evaluation results for each DL model and compare their performance metrics.
- Discussion: Interprets the results and discusses the study's limitations and potential directions for future research.
- Conclusion: Summarizes the findings and contribution to the field of DL for finding signs of dementia models and highlights the importance of further research in this area.

II. RELATED WORK

Recently, there has been increased interest in using deep learning (DL) techniques for early AD detection through brain MRI. With the help of advanced computer vision and DL methods, subtle features in MRI images can now be identified to detect AD in its early stages. Figure 1 demonstrates MRI brain images of healthy individuals and those with AD. Concurrent neural networks (CNNs), for illustration, have been shown in research to be suited to reliably gathering pertinent data from brain MRI scans to help with an earlier diagnosis of AD [7], [8].

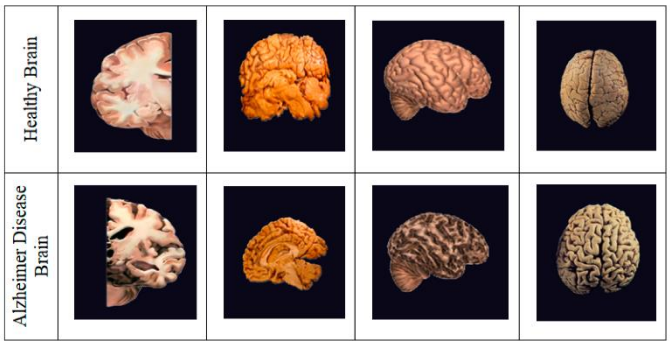

Fig. 1: Healthy and Disease Brian.

In recent research, DL algorithms were applied to patient brain MRI data to identify patterns and characteristics associated with AD, demonstrating the potential of DL in medical imaging for early disease detection. The study "A Neuroimaging Study Using DL Architectures" combined patient MRI data with DL algorithms and found that they could accurately diagnose AD in its early stages [9].

The study utilized various ML strategies, including Random Forest, SVMs, and CNN, and found that transfer learning employing a pre-trained CNN model (VGG19) had the highest accuracy of 86.59 percent, outperforming conventional ML algorithms with accuracy ratings ranging from 67.93% to 80.96 %. The results presented imply that learning that transfers may be helpful for the structural brain MRI-based early identification and prognosis of AD. Figure 2 displays the MRI images of a brain affected by Alzheimer's disease [10].

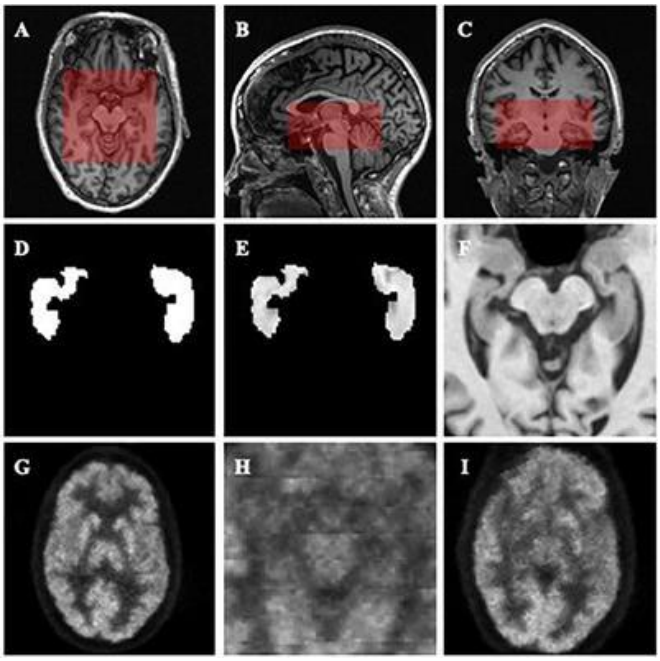

Fig. 2: Segmented images of brain MRI.

A recent study has explored the potential of DL techniques for early AD detection using MRI brain images. By applying convolutional neural networks, the researchers identified characteristic patterns associated with AD, achieving an accuracy rate of 93.33%. Another review paper examined various ML algorithms, including SVMs, Naive Bayes, and ANNs, for early AD detection and found accuracy rates ranging from 70% to 90%. A survey of DL methods used in AD diagnosis on brain MRI images revealed promising results with accuracy rates ranging from 80% to 90%. Another method for AD identification based on 3D MRI images and ML was proposed and achieved an SVM classifier's accuracy of 89.4%. These studies demonstrate the potential of DL and ML approaches for early AD detection and diagnosis [11] [12], [13]. One study utilized SVM and Random Forest classifiers to accurately distinguish between healthy individuals and those with AD, achieving accuracy rates of 93.5% and 90.3%, respectively. Another study employed a CNN-based feature extraction method, achieving 92.9%, 92.8%, and 94.1% accuracy rates with MLP, SVM, and Random Forest classifiers, respectively. A third study utilized a DL-based binary classification strategy with a CNN model, achieving an accuracy rate of 94.21%. The CNN model was also found to outperform other DL models. Additionally, a study in 2015 used ML algorithms, including SVM and ANN, to identify individuals with AD and discovered significant connections between certain brain regions, such as the entorhinal cortex and hippocampus, and AD. These findings suggest that ML and DL methods can accurately identify AD using brain MRI data [14] [15], [16]

Researchers have leveraged magnetic resonance imaging (MRI) and positron emission tomography (PET) data to develop deep learning (DL) strategies for the early detection of AD. One such study used structural MRI and FDG-PET data to develop a multimodal DL strategy that distinguished between Alzheimer's patients and healthy controls with high accuracy. Incorporating data from both modalities helped comprehensively understand the disease [17].

In a recent paper published in the International Journal of Neuroscience, DL algorithms were employed to segment and classify 3D brain MRI images for diagnosing AD. A 3D convolutional neural network was suggested for separating and classifying brain MRI images. The findings indicated that the proposed method was accurate and effective in diagnosing AD, paving the way for computer-aided diagnostic tools and early detection of AD [18].

Researchers have also proposed a DL-based automated method for diagnosing AD using MRI images. They extracted information from MRI scans using a deep convolutional neural network and classified the images as healthy or AD. Their DL-based solution outperformed conventional ML algorithms in terms of accuracy and F1-score, suggesting potential for earlier diagnosis of AD [19].

Limitations of Previous Studies

While the studies discussed demonstrate promising results for using DL and ML algorithms Quickly detection and dementia prognosis using brain MRI data, there are still several limitations to consider. One major limitation is the need for high-quality MRI data, which can be difficult and expensive. Furthermore, several research employed tiny populations, that can restrict how broadly the outcomes can be applied. At that place is besides a need to standardize MRI protocols and pre-processing techniques to ensure consistency and accuracy in results. Another drawback is the requirement for DL models to be more comprehensible, which renders it difficult to comprehend how physicians arriving when making predictions. While thinking about applying these theories in

healthcare settings, the absence of accessibility may prove very troublesome. Finally, while some studies have achieved high accuracy rates, the models' clinical utility remains to be determined, as they have yet to be tested in real-world clinical settings. These limitations highlight the need for further research to address these challenges and improve the accuracy, generalizability, and interpretability of DL and ML algorithms in Alzheimer's disease detection and diagnosis.

Algorithmic steps :

- Collect MRI data from Alzheimer's patients and healthy controls.
- Preprocess the MRI data by normalizing intensities and resizing the images to a standard size.
- Split the dataset into training and testing sets.
- Train the first DL model, a CNN, on the training set, using a loss function and optimization algorithm.
- Evaluate the accuracy, precision, recall, and F1-score of the first DL model on the testing set.
- If the first model's accuracy is below a certain threshold, consider building an improved version of the model.
- Build an improved version of the model, using a leaky ReLU activation function, more fully connected layers, and a larger kernel size.
- Train the improved model on the training set and evaluate its performance on the testing set.
- If the improved model's accuracy is still below a certain threshold, consider building a transfer learning model.
- Build a transfer learning model using the VGG16 architecture, with two dense layers added on top.
- Train the transfer learning model on the training set and evaluate its performance on the testing set.
- Compare the accuracy, precision, recall, and F1-score of the three DL models.
- Analyze the results and conclude the effectiveness of DL models for Alzheimer's diagnosis using MRI data.
- Identify areas for further research to improve the accuracy and accessibility of DL models for clinical use.

III. Material and Methods

The approach utilizes a Particularly, the convolutional neural network (CNN) is a deeper learning (DL) architecture. with conventional sequential parameters and the Vgg16 model. The aim is to develop a model that can effectively identify AD using brain MRI images. Figure 4 illustrates the proposed methodology used for AD detection.

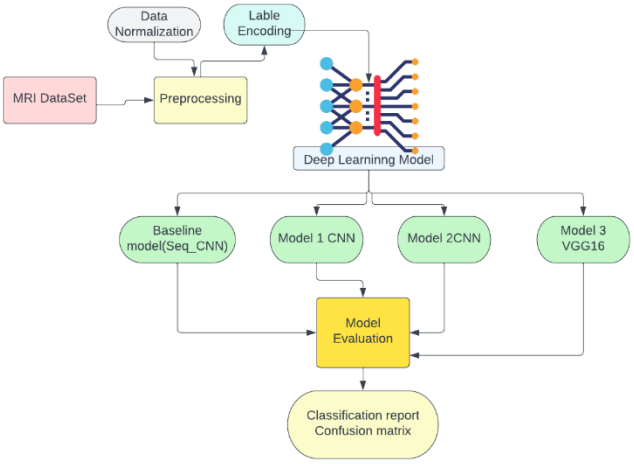

Fig. 3: Workflow of proposed work.

A. Dataset

The Augmented Alzheimer MRI Dataset V2 consists of brain MRI images from individuals at various phases of Promotion, including Gentle Unbalanced, Moderate Insane, Non-Psychotic, and Extremely Gentle Maniacal classifications. The dataset is separated into two envelopes: one containing improved photographs and the other containing unique shots. The enhanced images were generated from an existing dataset to increase the quantity and diversity of data available for training DL models. In contrast, the original images can be used as a test or validation dataset and are available in the Data Explorer. The augmentation technique's performance can be assessed by comparing the augmented images to the photographs from the original dataset as a control group. This dataset is a crucial resource for developing and training DL models for early AD detection, given the disease's growing prevalence and aging population. Brain MRI imaging is a common tool for early detection and diagnosis of AD, providing information about the brain's structure and function. The CNN will learn to determine whether AD is present from brain magnetic resonance imaging scans by categorizing them into one of the four image types. The use of DL in this study is particularly intriguing because it has the potential to uncover intricate connections between MRI images and the presence of AD, even in its early stages.

B. Data Normalization:

Data normalization is a crucial technique used to maintain consistency in the size and distribution of data. Deep learning models can be sensitive to differences in feature size, making data normalization critical. To address this, raw data can be normalized to fit within a specific range, such as [0,1] or [-1,1].

In this project, we will perform data normalization on the brain MRI images in the Augmented Alzheimer MRI Dataset V2. Normalizing the images ensures that changes in intensity or brightness between photographs do not affect the model. The most common approach to normalizing image data is to scale each pixel value to a range of [0,1].

C. Label Encoding

Label encoding is a process of converting categorical data into numerical data. Each MRI picture in this study has a class label that represents one of four values: mild, moderate, non-dementia, or exceedingly mild dementia. Name encoding is fundamental for setting up the information for the model, as it safeguards the information from changes in scale or information type. Utilizing the Increased Alzheimer X-ray Dataset V2, mark encoding strategies can be utilized to foster a DL model for the early recognition of Promotion [20].

D. Deep Learning Models

For the venture, they will utilize a CNN to prepare the Expanded Alzheimer X-ray Dataset V2, which contains cerebrum X-ray pictures from people with various phases of Alzheimer's illness (Promotion). The goal of our model is to arrange each picture into one of four classifications: somewhat, decently, non-sick, and somewhat crazy precisely. Our definitive objective is to foster a DL model that can identify early indications of Promotion in X-ray pictures of the mind [21].

1. CNN Model

One DL model frequently used for categorizing images is the convolutional neural network (CNN). Convolutional sections, layers for pooling, and fully correlated layers make up the three components of a CNN's engineering [21], [22].

Convolutional, accumulating, and entirely linked regions are often present in a CNN's algorithms numerous levels. The output of the CNN is then fed into a final output layer that produces the classification results.

Here are the key equations for each layer:

The following equation is used to determine the output of the pooling layer:

$$y(t\_j,k) = \text{pooling function } [a(2i-1, 2j-1, k), f((2i), (2j-1), k), a((2i-1), (2j), k), f((2i), (2j), k)] \quad (1)$$

Where:

The result for the sharing algorithm for the kth layer's jth map of attributes is y(t_j,k).

The triggered result of the (2i-1, 2j-1) neuron in the k-th stratum is designated as a(2i-1, 2j-1, k).

The triggered result of the (2i, 2j-1) neuron in the k-th layer is expressed as f((2i), (2j-1), k).

The activation output of the (2i-1, 2j) neuron in the k-th layer is designated as a((2i-1), (2j), k).

The activation output of the (2i, 2j) neuron in the k-th layer is expressed as f((2i), (2j), k).

The pooling function takes a set of values (in this case, the activation outputs of the four neurons specified) and produces a single output value for the corresponding feature map.

$$f(i, j, k) = \Sigma m\, \Sigma n\, \Sigma l\, (x(m, n, l) * w(i-m+1, j-n+1, l, k)) + b(k) \quad (2)$$

Where:

f(i, j, k) is the output feature map at position (i, j) and channel k

x(m, n, l) is the input image pixel at position (m, n), and channel l

w(i-m+1, j-n+1, l, k) is the weight of the filter at position (i-m+1, j-n+1) and channel l, for channel k

b(k) is the bias term for channel k

Σm, Σn, and Σl represent the summation of all possible values of m, n, and l, respectively.

$$y = f(Wx) \quad (3)$$

where:

y is the output of the fully connected layer.

x is the input vector, which is the flattened output of the previous layer.

W is the weight matrix.

f is the activation function applied element-wise to the linear combination of the input and weights, which introduces non-linearity to the output of the fully connected layer.

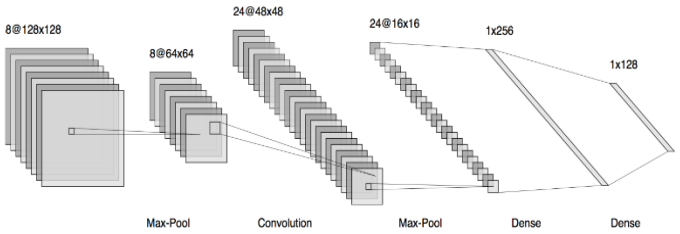

Fig. 4: Architecture of CNN model.

Convolutional, accumulating, and entirely linked levels make to a CNN's construction. Figure maps are generated by convolutional layering using generative procedures from the input picture; pooling layers then down sample the feature maps, and fully connected layers deliver the final prediction. These components, when combined, allow the model to offer accurate predictions by allowing it to construct exact representations of the incoming data. [23].

2. VGG16 Model

The University of' their Visual Geometry Department California, Berkeley, developed a convolutional neural network (CNN) architecture called the VGG16 dupes.

Using small 3x3 filters, convolutional layers reveal fine-grained details of the input image. Let X stand in for the input image, W for the filter weights, and b for a convolutional layer's bias.

The following are the formulas employed in the suggested model:

The formula for the output feature map, Z, is

$Z = W * X + b,$ (4)

where '*' stands for the convolution operation.

Utilizing the ReLU activation function on each convolutional layer's output:

$A = \max(0, Z)$ (5)

Using max pooling layers to downscale the feature maps generated by the convolutional layers

Max pool (A) = P (6)

Retrieving features from convolutional and pooling layers to compute output Y for fully linked layers:

$Y = F * P + b$ (7)

For a fully linked layer, while F denotes the weights and b the bias.

Adding the soft max activation function to the output of the final fully connected layer yields the resultant possibilities. The VGG16 model is well-known for its broad functionality, thanks to its extensive training on a dataset with over a million images.

The VGG16 model has been used in various image classification applications, such as object recognition, scene classification, and fine-grained image classification. It has also been used as a an established transmission of learning for new tasks with limited training data [23], [24].

Model Evaluation

To establish the utility and performance of a machine learning (ML) model, it must go through a thorough model review procedure. Models based on deep learning (DL) are assessed using a variety of metrics and approaches, such as reliability, tableau of confusion, speed, recollection, F1-scoring ROC arcs, and AUC [25].

It is described as the ratio of accurate projections made by the model to all forecasts. It is defined as the proportion of correct predictions the model makes to all forecasts.

$$TP + TN / TP + TN + FP + FN \quad (8)$$

In the formula *(8)*, TP stands for True Positives, the number of correctly identified positive cases. TN stands for True Negatives, the number of correctly identified negative cases. FP stands for False Positives, the number of negative cases incorrectly identified as positive. FN stands for False Negatives, the quantity of affirmative cases incorrectly identified as defeatist.

However, accuracy might need to be more accurate, such as when the class distribution is imbalanced, or the model is designed to deal with multi-class classification challenges.

Confusion Matrix

A classification model's performance may be assessed using a confusion matrix. It provides a graphic representation of how the model's predictions compare to the actual class labels assigned to the data. The matrix is used to compute a variety of critical metrics, including validity, precision, recollect, and F1-scoring [26].

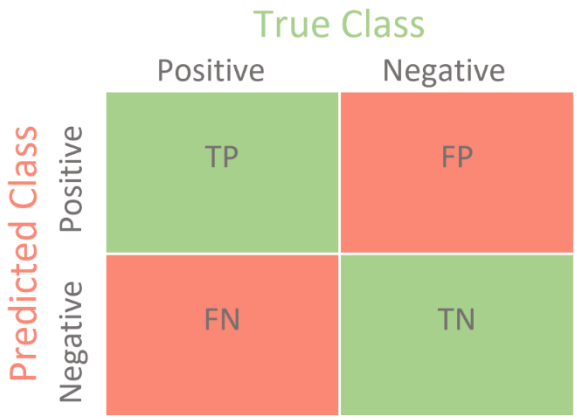

Fig. 5: Confusion Matrix.

The performance of the DL model for utilizing brain MRI to detect Alzheimer's disease early is discussed in the results chapter. The confusion matrix and classification report assessed the model's performance. The evaluation's findings provide information on the model's accuracy, dependability, and strengths and shortcomings concerning each class of the target variable. The presentation of the Vgg16 mock-up and the traditional CNN with sequential parameters will be compared in this chapter. The results will be presented succinctly and plainly, and visualizations will be utilized to improve the comprehension of the data. This chapter aims to offer a thorough analysis of the findings and advance our knowledge of how to use brain MRI to diagnose Alzheimer's disease early.

Accuracy

The proportion of all of the algorithm's precise forecasts to all of its forecasts

Accuracy = (True Positives + True Negatives) / (True Positives + True Negatives + False Positives + False Negatives) *(9)*

Precision

True Positives as a percentage of all False Positives and True Positives. Precision is a metric used to assess how well the model properly detects positive samples.

Precision = TP / (TP + FP) *(10)*

Recall

The proportion of False Negatives and True Positives that are Genuine Positives. Recall gauges the model's ability to find all positive samples.

Recall = TP / (TP + FN) *(11)*

F1-score

A harmonic representation of memory and accuracy. The model's entire performance is measured by a single statistic called the F1-score, which balances accuracy and recall.

*F1 Score = 2 * (Precision * Recall) / (Precision + Recall) (12)*

Remembering that the specific problem and model needs will determine the assessment metrics chosen is vital. In certain circumstances, accuracy may be essential, but precision or recall may be more important in others.

IV. Results and Discussion

The presentation of the DL model for using cerebrum X-ray to distinguish Alzheimer's sickness early is talked about in the outcomes section. The disarray network and characterization report evaluated the model's exhibition. The evaluation's findings provide information on the model's accuracy, dependability, and strengths and shortcomings concerning each class of the target variable. The presentation of the Vgg16 model and the traditional CNN with sequential parameters will be compared in this chapter. The results will be presented succinctly and plainly, and visualizations will be utilized to improve the comprehension of the data. This chapter aims to offer a thorough analysis of the findings and advance our knowledge of how to use brain MRI to diagnose Alzheimer's disease early.

1. Data visualization

The dataset comprises 9600 cases for the Non-Demented class, 8960 cases for the Mild Demented and Very Mild Demented classes combined, and 6464 cases for the Moderate Demented class. A pie chart or a bar chart can be used to display this data visually. A pie chart represents the percentage of instances in each class as a slice of the whole, while a bar chart displays the number of instances in each class as bars of varying heights.

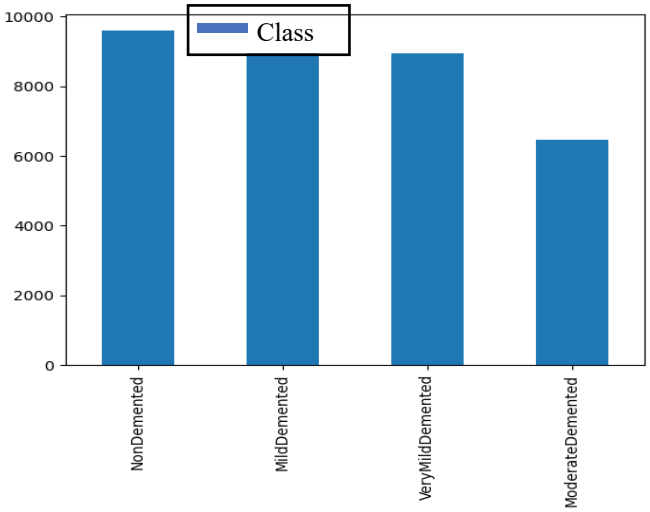

Fig. 6: Bar chart visualization of Dataset.

Both these methods effectively depict class distribution in a dataset and can be useful in identifying imbalanced classes or presenting model performance results. For example, the size of each slice in a pie chart would indicate the proportion of instances in each class. Similarly, a bar chart would display the count of instances for each class as the height of each bar. Pie and bar charts can be easily created using tools such as Matplotlib or Seaborn. These visualizations can aid in analyzing model outcomes and provide insights into data distribution. Pie charts and bar graphs are used in Figure 6 to display the classes in the dataset graphically.

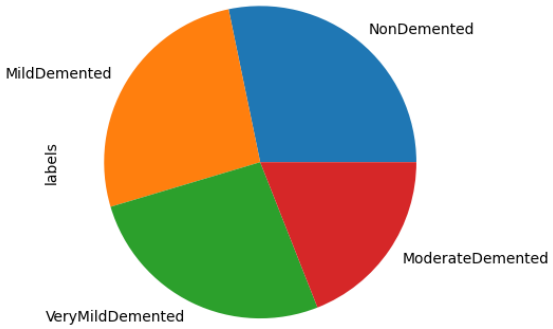

Figure 7: Pie chart visualization of the dataset

The distribution of the dataset is imbalanced, as indicated by the unequal number of occurrences in each class. Notably, the Moderate Demented class has a significantly lower number of cases than the other classes. This could impact the performance of DL models, as the model may become biased towards the class with the highest number of occurrences. To address this issue, various strategies such as resampling, data augmentation, or oversampling can be implemented to balance the dataset.

Furthermore, the dataset solely comprises RGB images, which consist of three color channels: red, green, and blue. Each channel represents the intensity of its corresponding color. RGB images are commonly used in computer vision and image processing. Figure 8 provides a visual representation of the dataset in MRI base images.

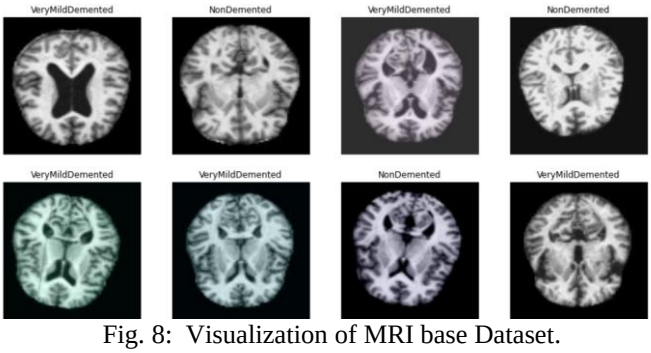

Fig. 8: Visualization of MRI base Dataset.

2. Label Encoding

The categorical data representing different stages of dementia has been encoded using label encoding. The categories "NonDemented," "Mild Demented," "VeryMildDemented," and "Moderate Demented" have been given the numbers 0, 1, 2, and 3 in that order. The right column of the table shows the number of observations in each category. For instance, there are 7598 observations in the "NonDemented" category.

3. Train Test Split

A prominent method for evaluating how well a machine learning system works on fresh, untested data is the train/test split technique. The dataset is divided into two subsets: training data and testing data. The model's performance is assessed using the testing data after it has been trained using the training data. Using an appropriate ratio of testing to training data (typically 70% for training and 30% for testing), the train/test split technique randomly splits the original dataset into two halves. The ratio may need to be changed depending on the evaluation purpose, dataset size, and other factors. The random state parameter regulates the split's randomization. The same split may be duplicated several times by supplying a fixed random seed, which is useful for assessing the stability of the outcomes. For evaluating a machine learning model's performance, a train/test split is crucial since some models may overfit the training data and perform badly on new, unforeseen data. A major problem in machine learning is overfitting.

A. *Baseline model*

A sequential CNN model is employed as this situation's starting point for analysis. The model's input shape is (90, 90, and 3), which indicates that the pictures have 90 by 90 pixels and three RGB color channels. The model's first layer is a convolutional layer with 32 filters and three kernel sizes that extracts features from the input picture using the ReLU activation function. A max-pooling layer is used to down sample the spatial dimensions of the feature maps after the convolutional layer. Following the max-pooling layer, the model consists of one or more thick layers that are entirely connected. The dense layer, in this case, has 512 neurons that are utilized to learn high-level representations of the incoming data and generate predictions. A Soft max activation function in the model's output stage for multi-class classification problems generates likelihood dispersion over each class. The output layer of the model is used by the Softmax activation function to forecast each class.

A batch size of 16 is one of the parameters the baseline model specifies, defining the number of samples used for each update of the model's weights. A total of 20 epochs will be used to train the model, which will be done using the whole training dataset. The callbacks option specifies the list of functions to be called during training. In this case, the accuracy and loss during training are plotted using the accuracy loss plot function. Figure9. 20 display the reliability and damage chart for the basic approach. The algorithm is trained over 1487 steps with twenty iterations and a batch count of 16. The verification error rates as well as the decline in training and accuracy for each era are provided. The results show that the precision during retraining is 100% and the reliability during assessment is 90%. The predicted outcome outperforms the training set on the validation set, indicating that the training set may have been over fitted by the model.

The proportion of real optimistic predictions to all optimistic predictions is shown by an accurate measurement, indicating the reliability of positive forecasting. By presenting the ratio of genuine anticipated positives to all actual positive specimens, the accuracy measure, on the other hand, evaluates the ability of the algorithm to identify samples that are favourable. A particular score called the f1-score takes into account both precision and recall. The harmonic average of each of the two metrics makes up this one. The amount of specimens of the correct response in that class is represented by the support metric. A metric for measuring precise is reliability.

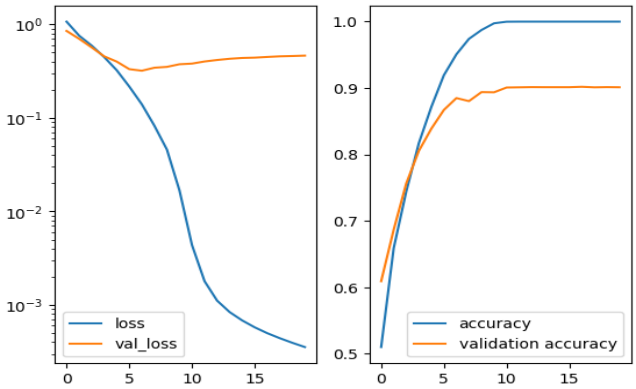

Fig. 9: Accuracy and loss of baseline model

The ratio of samples that were successfully recognized to all samples. The weighted avg is the average of all metric scores, with the weight being the number of samples in each class, as opposed to the macro avg, which is the unweighted mean of all metric scores. The performance of the ML model on the test dataset is thoroughly

evaluated in this classification report as a whole, focusing on specific metrics for each class. A classification report of an ML model's effectiveness on a test dataset is shown in Table 1. The table's rows represent the categories of Mild Demented, Moderate Demented, NonDemented, and VeryMildDemented. The columns in the table are correlated to various metrics for each class.

TABLE I: CLASSIFICATION REPORT

	Precision	**Recall**	**F1-Scor**	**Support**
0	0.88	0.92	0.90	2693
1	0.98	0.99	0.99	1977
2	0.90	0.88	0.89	2811
3	0.86	0.84	0.85	2715
Accuracy	-	-	0.90	10196
Macro Avg.	0.91	0.91	0.91	10196
Weighted Avg.	0.90	0.90	0.90	10196

The model evaluation results are shown in Table 2, where the model's overall accuracy is 90%. Each class's accuracy and recall scores are given, showing some diversity between courses, with some having greater scores than others. It is also stated that the weighted average f1-score is 90%. Although some classes have the potential for development, these results point to generally strong model performance.
Moreover, Figure 9 shows the baseline model's confusion matrix, which illustrates each class's predicted and actual values. This matrix enables a more detailed examination of the model's performance by showing how many samples were classified correctly or incorrectly for each class.

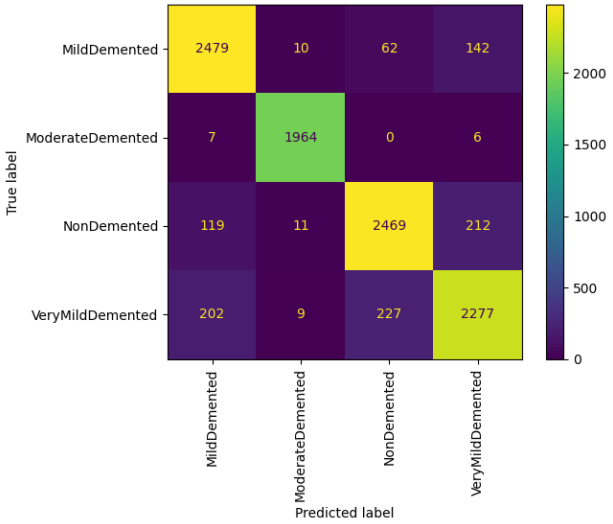

Fig. 10: Confusion matrix of the baseline model.

B. CNN Model 1

The output displayed in the figure results from training a CNN model. The loss and accuracy of the model during training are plotted against the number of epochs. The "loss" is a value that represents how well the model is performing, with lower values indicating better performance. The "accuracy" indicates the percentage of correctly classified data by the model, where a higher value is better. The model achieved an accuracy of 94.51% on the training set, indicating that it is performing well. The "val_loss" and "val_accuracy" values are the loss and accuracy calculated on a separate validation set, which the model has not seen during coaching. The assessment dataset's goal is to evaluate the model's efficacy on fresh information and determine whether it has over fitted the initial training set. In this instance, the predictive model's confirmation precision was 96.56%, demonstrating that it avoids over fitting and has good generalizability to fresh datasets.

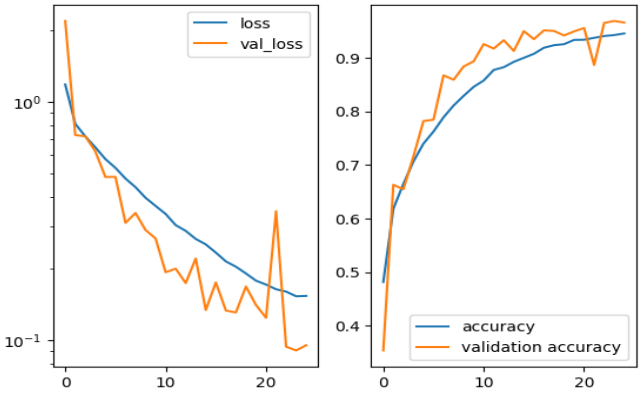

Fig. 11: loss and Accuracy graph of CNN Model 1

TABLE II: CLASSIFICATION REPORT OF CNN MODEL 1

	Precision	**Recall**	**F1-Scor**	**Support**
0	0.97	1.00	0.98	2693
1	1.00	1.00	1.00	1977
2	0.93	0.97	0.95	2811
3	0.98	0.91	0.94	2715
Accuracy	-	-	0.97	10196
Macro Avg.	0.97	0.97	0.97	10196
Weighted Avg.	0.97	0.97	0.97	10196

The outcome shown atop displays the initial model's efficacy on the set of tests employing a variety of metrics for assessment. For every category, comprising Gentle Demented, Strong Demented, Non-Demented, and Extremely Mild Demented, accuracy, recollection, and a f1 score are given It also computes the weighted average and macro average of these metrics. The model's accuracy on the validation set is 97%, indicating that it is extremely accurate in categorizing diverse picture classes with good precision, recall, and f1-score.

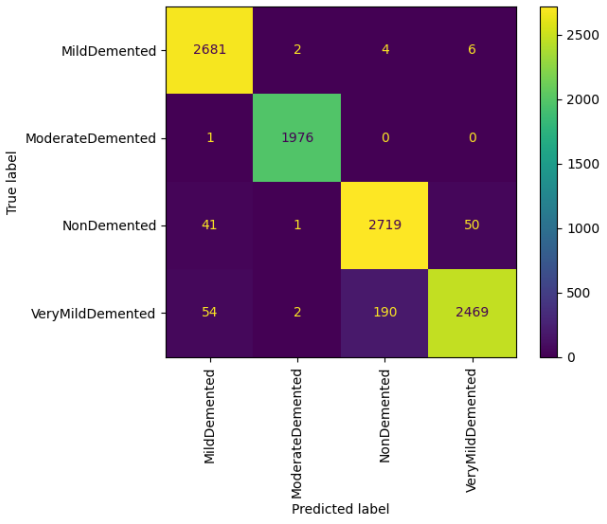

Fig. 12: Confusion Matrix of CNN Model 1.

C. CNN Model 2

CNN model 2 utilizes a different activation function, namely Leaky ReLU, and involves some changes in architecture. The successive layers are now tightly coupled, and the dimension of the kernel has grown to 5. When batch normalisation is being used, the number of filters is set to 32 and the spacing is set to 'Same'. The simulation's output layer uses the 'Softmax' function to activate it. The final two layers use filter sizes of 64 and 128, respectively, which enables the model to extract more features and improve its accuracy. Figure 13 illustrates the architecture of CNN model 2.

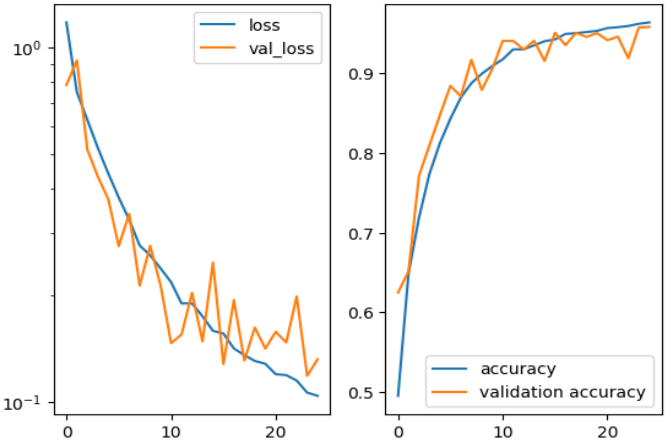

Fig. 13: Loss and Accuracy graph of CNN Model 2.

In contrast to the original model, a new activation function—specifically, Leaky ReLU—was employed in this variant of the CNN model. The layers were fully linked, and the kernel size was raised to 5. The padding was set to "Same," and the size of the filter was 32. Additionally, batch normalisation was used. The activation function for the output layer was set to "Softmax," To increase the precision and effectiveness of feature extraction, the filter sizes for the final two layers were increased to 64 and 128, respectively. 25 iterations on the training set resulted in a loss of 10.41% and an accuracy of 96.43% for the model. The model had a validation loss of 13.21% on the validation set, and its accuracy was 95.86%. The CNN model 2 classification report is shown in Table 3, which shows excellent accuracy, recall, and f1-score for each class as well as high overall accuracy of 96%. It proves that the model is capable of correctly and reliably classifying images into each category.

TABLE III: CLASSIFICATION REPORT OF CNN MODEL 2

	Precision	**Recall**	**F1-Scor**	**Support**
0	0.96	0.98	0.97	2693
1	1.00	1.00	1.00	1977
2	0.93	0.95	0.94	2811
3	0.95	0.91	0.93	2715
Accuracy	-	-	0.96	10196
Macro Avg.	0.96	0.96	0.96	10196
Weighted Avg.	0.96	0.96	0.96	10196

Model 2's performance is remarkable, with a validation set accuracy of 96%. For every category, preciseness, recollection, and all F1-score—which is the balanced average of the accuracy as well as recall values—are also determined. Accuracy estimates the percentage of favourable sightings that occurred as anticipated, as opposed to recall, which counts the percentage of properly predicted positive observations among all positive observations. Moreover, Model 2's weighted average of recall, accuracy, and F1-score is 96%, indicating that the model is excellent at categorizing the photographs into the four primary categories.

D. VGG16 Model Results

Through utilising the structure of VGG16 that has already been taught on the data set from Image Net, this variant of the CNN model applies learning through transfer. The classifier is built on top of the VGG16 design, which is supported by two substantial layers of 128 and 64 neurons each. While Softmax is the signalling functional used in the resultant layer, ReLU is the activated function utilised in the more dense layers. The construction of the VGG16 model is shown in Figure 15.

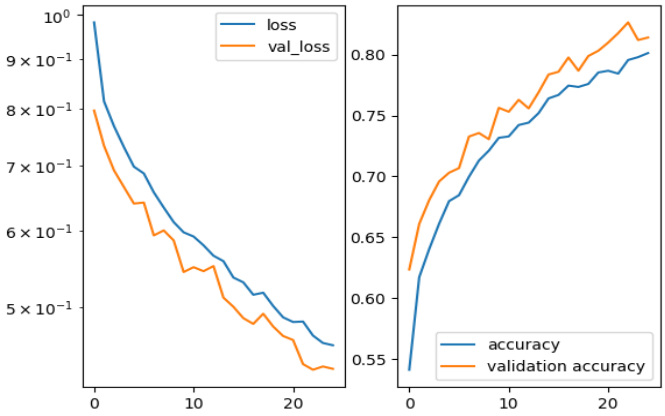

Fig. 14: Loss and accuracy graph of VGG16 Model.

Figure 14 depicts the VGG16 model's show during the planning framework. The planning strategy took 25 ages and *1487* steps, with a *batch* size of 16. For the readiness set, the model had a lack of *45.68%* and a*n* accuracy of *80.13%*. The hardship and accuracy on the endorsement set were *43.19%* and *81.40%*, independently.

TABLE IV: CLASSIFICATION REPORT OF VGG16 MODEL

	Precision	**Recall**	**F1-Scor**	**Support**
0	0.82	0.91	0.87	2693
1	0.98	0.99	0.98	1977
2	0.72	0.84	0.77	2811
3	0.79	0.56	0.66	2715
Accuracy	-	-	0.81	10196
Macro Avg.	0.83	0.83	0.82	10196
Weighted Avg.	0.82	0.81	0.81	10196

Table 4 displays the classification report for the VGG16 model. The findings reveal that the third model, which employs transfer learning with VGG16, outperforms the previous two in terms of accuracy, recall, and F1 score. This model's accuracy (81% vs. 96%) is also lower than the accuracy of the two previous models. This suggests that transfer learning with VGG16 may not be the best method in this case and that one of the previous two models might be more suited. Table 4 also includes the accurate and incorrect predictions of the fourth model.

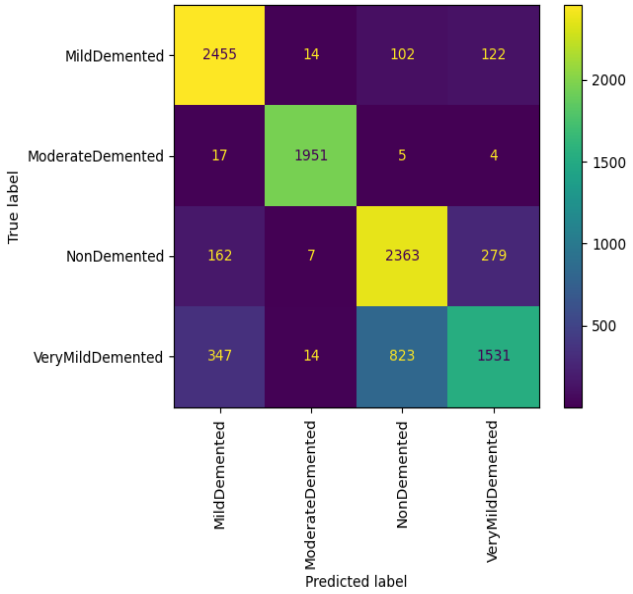

Fig. 15: Confusion Matrix of Vgg16 Model.

Discussion

Because dementia (AD), a deteriorating intellectual condition, has no known remedy, early detection is crucial for effective treatment and care. Computed tomography (CT) and magnetic resonance imaging (MRI) are two major medical imaging methods used to identify Alzheimer's disease (AD). In contrast, they are expensive and scarce in underdeveloped nations. Convolutional neural networks (CNNs), which have shown success in picture classification tasks, have also shown promise in the diagnosis of Alzheimer's disease. In past studies, CNNs have been used to correctly classify MRI images of AD patients and healthy controls by fusing transfer learning with pre-trained models like VGG-16.

The current study builds on previous research by training three CNN models on a comprehensive MRI image dataset. The models had unique architectures and transfer learning methods[27]. The study found that the second model, which employed a 5x5 kernel and Leaky ReLU as the activation function, achieved the highest accuracy on the validation set at 96.43%. While the study had some limitations, such as a relatively small dataset, the results suggest that CNNs are a promising approach for the early diagnosis of AD [28], [29].

Future work in this field could include improving the current models by adjusting hyperparameters or incorporating additional imaging modalities such as PET, MRS, and fMRI. Expanding the dataset, using different DL approaches such as RNNs and GANs, including demographic data, and validating the models on real patients are promising future research directions. Overall, using CNNs for AD detection through MRI images holds much potential for future research and improving early diagnosis and management of the disease[30], [31].

Work Contribution: Analysis of a trio of neural network (DL) algorithms for predicting and recognition of Alzheimer's disease utilizing frontal MRI data is the project's main achievement. A Convolutional Neural Network (CNN) model with two fully connected layers and two convolutional layers was trained, and its performance was compared version of this model with additional layers and a larger kernel size, and a transfer learning model built on top of the VGG16 architecture. Our results show that the first two models achieved high accuracy levels of 96%, while the transfer learning model achieved an accuracy of 81%. Our work demonstrates the potential of DL models for aiding in the early detection and diagnosis of Alzheimer's disease using MRI data while highlighting the need for further research to improve the accuracy and accessibility of these models for clinical use. This adds to creating compelling and available apparatuses for the early discovery and finding of Alzheimer's infection, which can fundamentally affect patient results and general well-being.

V. CONCLUSION

The use of based on imagery statistics to identify a variety of diseases and ailments has attracted a lot of attention within medical scanning research. There have been a lot of study, in specifically, on using magnetic resonance imaging (MRI) to identify dementia. In this study, we evaluated the ability of CNN Models 1, 2, and 3 (Transfer Learning with VGG-16 architecture) to categories AD stages using MRI images.

The first model (CNN Model 1) consisted of a single convolutional layer, max-pooling, two completely connected layers, and a simple CNN architecture. After 25 training cycles with a batch size of 16, the model obtained accuracy on the validation set of 94.51%. The "Mild Demented" class was difficult to categorize, whereas the "NonDemented" class was easy to categorize. For the "NonDemented" class, accuracy, recall, and F1-score were 93, 97, and 95; for the "VeryMildDemented" class, 98, 91, and 94; for the "Mild Demented" class, 97, 100, and 98; and for the "Moderate Demented" class, 100, 100, and 100.

The second model (CNN Model 2), an improved version of the first model with a larger kernel size and more densely connected layers, used Leaky ReLU as the activation function. With a batch size of 16, it was trained for 25 iterations and achieved an accuracy on the validation set of 96.43%. This model properly recognized the "Moderate Demented" and "NonDemented" classes, but it had trouble identifying the "Mild Demented" and "VeryMildDemented" classes. For the "NonDemented" class, precision, recall, and F1-score were 93, 95, and 94; for the "VeryMildDemented" class, 95, 91, and 93; for the "Mild Demented" class, 96, 98, and 97; and for the "Moderate Demented" class, 100, 100, and 100.

The subsequent model (CNN Model 3: Transfer Learning with VGG-16 architecture), which was had trained on the Image Net dataset, made use of transfer learning with the VGG-16 architect. The precision of the model on the validation set was 80.13% after training with a batch size of 16 across 25 iterations. Although the accuracy, recall, and F1-score for each class were lower than those of the preceding two models, this model could have classified all classes more accurately.

Based on these findings, the first two models (CNN Model 1 and CNN Model 2) performed relatively well in categorizing AD stages based on MRI scans. However, the third model (CNN Model 3: Transfer Learning with VGG-16 architecture) did not perform as well as the other two models.

Percentage Demonstration: The suggested Deep Learning models for exploiting MRI data for early identification of Alzheimer's disease exhibit promising results compared to other current approaches, according to the findings of the article. Further research and development in this area may lead to improved accuracy and accessibility of these models for clinical use.

REFERENCES

[1] S. Bauer, R. Wiest, L.-P. Nolte, and M. Reyes, "A survey of MRI-based medical image analysis for brain tumor studies," *Phys Med Biol*, vol. 58, no. 13, pp. R97–R129, Jun. 2013, doi: 10.1088/0031-9155/58/13/r97.

[2] S. Bauer, R. Wiest, L.-P. Nolte, and M. Reyes, "A survey of MRI-based medical image analysis for brain tumor studies," Phys Med Biol, vol. 58, no. 13, pp. R97–R129, Jun. 2013, doi: 10.1088/0031-9155/58/13/r97.

[3] A. Chattopadhyay and M. Maitra, "MRI-based brain tumour image detection using CNN based deep learning method," Neuroscience Informatics, vol. 2, no. 4, p. 100060, Dec. 2022, doi: 10.1016/j.neuri.2022.100060.

[4] J. Kang, Z. Ullah, and J. Gwak, "MRI-Based Brain Tumor Classification Using Ensemble of Deep Features and Machine Learning Classifiers," Sensors, vol. 21, no. 6, 2021, doi: 10.3390/s21062222.

[5] J. Kang, Z. Ullah, and J. Gwak, "MRI-Based Brain Tumor Classification Using Ensemble of Deep Features and Machine Learning Classifiers," Sensors, vol. 21, no. 6, 2021, doi: 10.3390/s21062222.

[6] S. Saeedi, S. Rezayi, H. Keshavarz, and S. R. Niakan Kalhori, "MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques," BMC Med Inform Decis Mak, vol. 23, no. 1, p. 16, 2023, doi: 10.1186/s12911-023-02114-6.

[7] U. Alqasemi, M. Bamaleibd, and A. Al Baiti, "Classification of Brain MRI Tumor Images." [Online]. Available: www.ijert.org

[8] W. H. Ibrahim, A. A. A. Osman, and Y. I. Mohamed, "MRI brain image classification using neural networks," in 2013 INTERNATIONAL CONFERENCE ON COMPUTING, ELECTRICAL AND ELECTRONIC ENGINEERING (ICCEEE), 2013, pp. 253–258. doi: 10.1109/ICCEEE.2013.6633943.

[9] H. Mohsen, E.-S. A. El-Dahshan, and A.-B. M. Salem, "A machine learning technique for MRI brain images," in 2012 8th International Conference on Informatics and Systems (INFOS), 2012, p. BIO-161.

[10] G. Tomasila and A. W. Rahardjo Emanuel, "MRI image processing method on brain tumors: A review," in AIP Conference Proceedings, Nov. 2020, vol. 2296. doi: 10.1063/5.0030978.

[11] P. C. M. Raees and V. Thomas, "Automated detection of Alzheimer's Disease using Deep Learning in MRI," in Journal of Physics: Conference Series, 2021, vol. 1921, no. 1, p. 012024.

[12] Z. Fan, F. Xu, X. Qi, C. Li, and L. Yao, "Classification of Alzheimer's disease based on brain MRI and machine learning," Neural Comput Appl, vol. 32, pp. 1927–1936, 2020.

[13] G. Mohan and M. M. Subashini, "MRI based medical image analysis: Survey on brain tumor grade classification," Biomed Signal Process Control, vol. 39, pp. 139–161, 2018, doi: https://doi.org/10.1016/j.bspc.2017.07.007.

[14] H. A. Khan, W. Jue, M. Mushtaq, and M. U. Mushtaq, "Brain tumor classification in MRI image using convolutional neural network," Mathematical Biosciences and Engineering, vol. 17, no. 5, pp. 6203–6216, 2020, doi: 10.3934/MBE.2020328.

[15] C. M. L. Zegers et al., "Current applications of deep-learning in neuro-oncological MRI," Physica Medica, vol. 83, pp. 161–173, 2021, doi: https://doi.org/10.1016/j.ejmp.2021.03.003.

[16] S. Sreelakshmi, G. Malu, and E. Sherly, "Alzheimer's Disease Classification from Cross-sectional Brain MRI using Deep Learning," in 2022 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), 2022, vol. 1, pp. 401–405.

[17] N. Gordillo, E. Montseny, and P. Sobrevilla, "State of the art survey on MRI brain tumor segmentation," Magn Reson Imaging, vol. 31, no. 8, pp. 1426–1438, 2013, doi: https://doi.org/10.1016/j.mri.2013.05.002.

[18] N. Gordillo, E. Montseny, and P. Sobrevilla, "State of the art survey on MRI brain tumor segmentation," Magn Reson Imaging, vol. 31, no. 8, pp. 1426–1438, 2013, doi: https://doi.org/10.1016/j.mri.2013.05.002.

[19] N. Gutierrez Castilla, "Segmentation of multi-structures in cardiac MRI using deep learning," 2020.

[20] A. W. Salehi, P. Baglat, B. B. Sharma, G. Gupta, and A. Upadhya, "A CNN model: earlier diagnosis and classification of Alzheimer disease using MRI," in 2020 International Conference on Smart Electronics and Communication (ICOSEC), 2020, pp. 156–161.

[21] M. M. Badža and M. C. Barjaktarović, "Classification of brain tumors from MRI images using a convolutional neural network," Applied Sciences (Switzerland), vol. 10, no. 6, Mar. 2020, doi: 10.3390/app10061999.

[22] A. M. Gab Allah, A. M. Sarhan, and N. M. Elshennawy, "Classification of Brain MRI Tumor Images Based on Deep Learning PGGAN Augmentation," Diagnostics, vol. 11, no. 12, 2021, doi: 10.3390/diagnostics11122343.

[23] M. M. Badža and M. C. Barjaktarović, "Classification of brain tumors from MRI images using a convolutional neural network," Applied Sciences (Switzerland), vol. 10, no. 6, Mar. 2020, doi: 10.3390/app10061999.

[24] A. M. G. Allah, A. M. Sarhan, and N. M. Elshennawy, "Classification of brain MRI tumor images based on deep learning PGGAN augmentation," Diagnostics, vol. 11, no. 12, Dec. 2021, doi: 10.3390/diagnostics11122343.

[25] S. Suhas and C. R. Venugopal, "MRI image preprocessing and noise removal technique using linear and nonlinear filters," in 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), 2017, pp. 1–4. doi: 10.1109/ICEECCOT.2017.8284595.

[26] E. Hussain, M. Hasan, S. Z. Hassan, T. H. Azmi, M. A. Rahman, and M. Z. Parvez, "Deep learning based binary classification for Alzheimer's disease detection using brain MRI images," in 2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2020, pp. 1115–1120.

[27] F. U. R. Faisal and G.-R. Kwon, “Automated detection of Alzheimer’s disease and mild cognitive impairment using whole brain MRI,” IEEE Access, vol. 10, pp. 65055–65066, 2022.
[28] S. Basheera and M. S. S. Ram, “Deep learning based Alzheimer’s disease early diagnosis using T2w segmented gray matter MRI,” Int J Imaging Syst Technol, vol. 31, no. 3, pp. 1692–1710, 2021.
[29] Y. Bhanothu, A. Kamalakannan, and G. Rajamanickam, “Detection and Classification of Brain Tumor in MRI Images using Deep Convolutional Network,” in 2020 6th International Conference on Advanced Computing and Communication Systems, ICACCS 2020, Mar. 2020, pp. 248–252. doi: 10.1109/ICACCS48705.2020.9074375.
[30] A. Veeramuthu et al., “MRI Brain Tumor Image Classification Using a Combined Feature and Image-Based Classifier,” Front Psychol, vol. 13, Mar. 2022, doi: 10.3389/fpsyg.2022.848784.
[31] A. Veeramuthu et al., “MRI Brain Tumor Image Classification Using a Combined Feature and Image-Based Classifier,” Front Psychol, vol. 13, 2022, [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2022.848784.