Comparative Analysis of Deep Learning Models for Multiclass Alzheimer’s Disease Classification

Raghav Agarwal¹, Abbaraju Sai Sathwik², Deepthi Godavarthi³* and Janjhyam Venkata Naga Ramesh⁴

¹School of Computer Science & Engineering (SCOPE), VIT-AP University, Amaravati, Andhra Pradesh, India
²School of Computer Science & Engineering (SCOPE), VIT-AP University, Amaravati, Andhra Pradesh, India
³School of Computer Science & Engineering (SCOPE), VIT-AP University, Amaravati, Andhra Pradesh, India
⁴Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur Dist., Andhra Pradesh - 522502, India

Abstract

INTRODUCTION: The terrible neurological condition is known worldwide; millions of individuals are affected with Alzheimer's disease (AD). Effective treatment and management of AD depend on early detection and a precise diagnosis. An effective method for identifying anatomical and functional abnormalities in the brain linked to AD is magnetic resonance imaging (MRI).

OBJECTIVES: However, manual MRI scan interpretation requires a lot of time and is inconsistent between observers. The automated analysis of MRI images for AD identification and diagnosis using deep learning techniques has shown promise.

METHODS: In this paper, we present a convolutional neural network (CNN)-based deep learning model for automatically classifying MRI images for Alzheimer’s (AD) and a healthy control group. A huge dataset of MRI scans was used to train the CNN, which distinguished between AD and healthy control groups with excellent accuracy.

RESULTS: Additionally, we looked into how transfer learning may be used to enhance pre-trained models and boost CNN performance. We discovered that transfer learning considerably increased the model’s accuracy and decreased overfitting. Our findings show that MRI scans may be used to precisely detect and diagnose AD utilizing approaches to deep learning and machine learning.

CONCLUSION: These techniques may improve the efficiency and accuracy of AD diagnosis and enable early disease identification, resulting in better AD management and therapy.

Keywords: Deep learning models, Multiclass classification, Comparative analysis, Transfer learning

1. Introduction

Alzheimer's disease (AD), a challenging and deadly fatal neurological disorder, affects millions of people worldwide. Progressive cognitive decline, memory loss, and changes in behavior and personality are the defining characteristics of AD [1,2]. A mix of clinical evaluations, cognitive tests, and imaging techniques, particularly magnetic resonance imaging (MRI), are used in the current Alzheimer's disease (AD) diagnostic standards. However, it can be difficult to correctly diagnose AD, and more effective and precise diagnostic techniques are required. Recently, potential methods for the interpretation of medical images, particularly MRI scans, to assist in the identification and diagnosis of AD have developed using deep learning and machine learning techniques [3]. Large datasets may be automatically analyzed and interpreted using these approaches, which also enable the detection of subtle changes in the structure and function of the brain that would not be apparent to the naked eye.

For the automated categorization of MRI scans into groups for AD and healthy controls, we describe a deep learning model...
model in this research that uses a convolutional neural network (CNN) as its foundation. A huge dataset of MRI scans was used to train the CNN, which distinguished between AD and healthy control groups with excellent accuracy [5]. Additionally, we looked into how transfer learning may enhance pre-trained models and boost CNN performance. Deep learning techniques might significantly improve the effectiveness and precision of AD diagnosis, which is essential for the disease's successful management and therapy. In this study, we demonstrate the viability and efficiency of using deep learning approaches to MRI scan-based AD detection and diagnosis.

Recent developments in biomedical imaging have been greatly facilitated by a relatively new DL paradigm. CNN is now the most often used DL architecture due to its notable performance in image analysis [6]. In contrast to standard ML, DL makes it possible for latent appearance to automatically generalize from low to high factor appearance. Thus, it's possible that DL uses less image preprocessing and requires less prior knowledge of other challenging processes [7,8], like choosing features, resulting in a much more objective and neutral approach.

In this study, for quick interpretation and diagnosis, a novel automated deep-learning approach is suggested in this work of brain MRI images. The study being presented main contributions are as follows:

i. A transfer learning diagnostic architecture that classifies Dementia is classified on MRI scans as Mild, Moderate, Non-Demented, or Very Mild.

ii. The analytical comparison of well-known techniques such as Xception, VGG models, ResNet models, MobileNet models, InceptionV3, and DenseNet models for the diagnosis of Alzheimer’s AD.

iii. Performance indicators including the ROC curve, F1-score, accuracy, precision, recall, and others have been used to assess how well the proposed system performs in comparison to rival models.

The following sections make up the remainder of the paper: Section 2 describes the literature survey, Section 3 outlines the methodology used in this study, the Results of the experimental data are covered in Section 4, and the conclusion and a suggestion for the future are provided in Section 5.

2. Literature Survey

In recent years, a great deal of investigations and research initiatives have focused on the identification of Alzheimer’s disease utilizing MRI. Several reliable works are evaluated in this section. Several DL-based techniques have recently been developed to detect Alzheimer’s disease on MRI images. To solve the issue of the original evolutionary algorithms' poor performance in feature selection, several academics have developed alternative methods to enhance their ability to categorize data. Even though it is common to use shared datasets, the results of various research have produced a range of conclusions. This is mostly because, even when the same process is employed, different parameters are used. The classification procedure and the existing models were altered in several experiments. To improve accuracy rates, this was done.

<table>
<thead>
<tr>
<th>S. No</th>
<th>Paper Title</th>
<th>Computational Models</th>
<th>Accuracy</th>
</tr>
</thead>
</table>

[1] composed a comparison between different transfer learning models including VGG models and ResNet models. Out of all the models used in the study, the model with the best performance, the ResNet101, had the highest accuracy about 99.51%.

[2] compared a custom CNN with a fine-tuned EfficientNetV2 on only five photos, resulting in getting an accuracy of 100% in custom CNN by outperforming EfficientNetV2 with 95%.

[3] illustrated several transfer learning and machine learning models and got the highest accuracy of 75.17% with ResNet50 as mentioned in Table 1.

3. Methodology

The proposed initiative aims to develop and evaluate employing magnetic resonance imaging (MRI) data, and deep learning models for the identification and diagnosis of...
Comparative Analysis of Deep Learning Models for Multiclass Alzheimer’s Disease Classification

Alzheimer’s disease (AD). The following are the precise goals of this work.

3.1. Data Gathering

We gathered a dataset of MRI scans from Kaggle [4]. Keeping class domination in mind 6400 MRI images make up the dataset, which has been split into three folders with subfolders for each class i.e., Mild Demented (896 images), non-Demented (3200 images), Very Mild Demented (2240 images), Moderate Demented (64 images), with 80% of the training images, 10% for validation and 10% of the test images.

The dataset was pre-processed to eliminate noise, motion artifacts, and other picture distortions and was made up of T1-weighted images. To standardize the image acquisition settings, bias correction, and skull stripping were also applied to the pictures.

3.2. Feature Extraction

We used an approach based on deep learning to extract characteristics from the MRI scans [9]. We employed a convolutional neural network (CNN) architecture that has already been trained, such as Xception, VGG Family, ResNet Family, MobileNet Family, InceptionV3, and DenseNet Family, to extract characteristics from the photographs. Before entering a classifier, the characteristics that were retrieved were flattened.

3.3. Model Development

For the categorization of MRI scans into AD and healthy control groups, we created several models. We employed transfer learning techniques including Xception, VGG, Resnet, MobileNet, and DenseNet families. We used a combination of training and validation datasets to train our models, and we fine-tuned the hyperparameters to enhance the models’ performance.

3.4. Model Evaluation

Using common measures, we evaluated the performance of our models using metrics such as accuracy, precision, recall, F1-score, and area under the curve (AUC) on the test dataset [10,11]. Additionally, we evaluated how well our models performed in comparison to other cutting-edge techniques.

The suggested approach may help with early disease identification and increase the precision and effectiveness of AD diagnosis utilizing MRI images. The deep learning used in this work may also aid in the discovery of novel AD biomarkers and the advancement of our knowledge of the pathophysiology of the illness.

3.5. Proposed Classification Methods and Techniques

The identification and diagnosis of Alzheimer’s disease (AD) are four-class image classification tasks where deep learning has demonstrated promising outcomes. For the categorization of MRI scans into groups representing AD and healthy control subjects [12,13], we propose to investigate several deep learning techniques in this study. Fig.1 demonstrates our applied strategy.

**CNNs (convolutional neural networks)**

Among deep learning algorithms, CNNs are exceptional at classifying images. We will investigate several CNN architectures for the categorization of MRI scans into AD and healthy control groups, including VGG16, ResNet50, and InceptionV3. To maximize performance, we shall tweak the models’ hyperparameters as mentioned in Table 2.

<table>
<thead>
<tr>
<th>Hyper-parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image size</td>
<td>224 x 224</td>
</tr>
<tr>
<td>Weight</td>
<td>ImageNet</td>
</tr>
<tr>
<td>Epochs</td>
<td>100</td>
</tr>
<tr>
<td>Batch size</td>
<td>32</td>
</tr>
<tr>
<td>Optimizer</td>
<td>Adam</td>
</tr>
<tr>
<td>Learning rate</td>
<td>0.001</td>
</tr>
<tr>
<td>Loss</td>
<td>Categorical cross-entropy</td>
</tr>
</tbody>
</table>

### Table 2. Hyperparameter for models

**Transfer Learning**

Transfer learning is a method for enhancing the performance of models on smaller datasets by employing pre-trained models on big datasets. We will look into ways to enhance classification accuracy utilizing transfer learning with pre-trained CNN models and extract features from the MRI images.

- **Xception.** The Xception model has displayed outstanding performance in a variety of applications for photo classification, such as medical image analysis. By
utilizing depthwise separable convolutions, the model's capacity to generalize to new data is improved while the number of parameters is reduced.

- **VGG-16.** In this work, we classified MRI scans for multiclass AD. The model's hyperparameters will be adjusted, and its performance will be assessed using common metrics like precision, area under the curve (AUC), recall, and accuracy. In the model's training, validation, and testing, we'll utilize the Kaggle dataset, which consists of MRI scans from AD patients and healthy controls.

  In general, we anticipate that the VGG16 model will succeed in classifying MRI scans into different classes of AD with high accuracy. The model has performed admirable in several picture classification tasks, and we feel that it can be adjusted to our job with some minor input and output layers.

- **VGG-19.** The input layer of the VGG19 model will be changed to take the MRI images as input to make it suitable for our job. We will also add a new softmax layer with four nodes that correspond to the different classes of AD instead of the model's output layer. To focus on our objective, we will fine-tune the weights of the remaining layers while freezing the weights of the model's initial few layers, which are in charge of feature extraction. The VGG19 model contains more layers than the VGG16 model, which may enable it to record MRI images with more complicated characteristics.

- **ResNet.** ResNet50, ResNet50V2, ResNet101, ResNet101V2, ResNet152, and ResNet152V2 were the six distinct ResNet models we employed in our research. Deep convolutional neural network models called ResNet models have been demonstrated to be effective in image categorization applications. The skip connections in the ResNet models enable the models to learn the residual functions and facilitate model optimization.

  In comparison to the other ResNet models, the ResNet50 model contains 50 layers and is a rather shallow model. ResNet101 is a deeper model than ResNet50 and comprises 101 layers. The deepest ResNet model is ResNet152, which contains 152 layers. The ImageNet dataset serves as a solid start for transfer learning to our objective of classifying Alzheimer's disease, and the ResNet models are pre-trained on this dataset.

- **MobileNet.** In our research, we utilized the MobileNet and MobileNetV2 models. Lightweight convolutional neural networks called MobileNet models were created primarily for mobile and embedded vision applications. To minimize the complexity of the models' computations and several parameters while retaining high accuracy, MobileNet models employ depthwise separable convolutions.

  Compared to other deep learning models, the MobileNet model is quite shallow with only 28 layers. In comparison to MobileNet, the MobileNetV2 model is deeper and more complicated, with 53 layers. The ImageNet dataset was used to pre-train both the MobileNet and MobileNetV2 models, which serves as a useful initialization for transfer learning to our objective of classifying Alzheimer's disease.

- **DenseNet.** DenseNet121, DenseNet169, and DenseNet201 were the three distinct DenseNet models we employed in our research. Convolutional neural networks called DenseNet models were created expressly to address the vanishing gradient issue in deep networks. In DenseNet models, all layers are connected directly to one another via skip connections, allowing for improved information flow across the network and better gradient flow during backpropagation.

  There are 121 layers in the DenseNet121 model, 169 layers in the DenseNet169 model, and 201 layers in the DenseNet201 model. With an increase in layers, models get more complicated and have more layers. All DenseNet models have already been trained on the ImageNet dataset, which gives transfer learning a suitable starting point for our objective of classifying Alzheimer's disease.

- **InceptionV3.** Convolutional neural network architecture called the InceptionV3 model was first put out within the context of the ImageNet Large-Scale Visual Recognition Challenge. From the incoming data, complex properties should be extracted, the InceptionV3 model combines convolutional layers with various filter sizes and max pooling layers.

  The ImageNet dataset, which contains more than one million labeled pictures from 1000 categories, served as the pre-training data for the 48 convolutional layers of the InceptionV3 model. In this research, we improved the InceptionV3 model using the MRI dataset to classify Alzheimer's illness.
4. Results & Discussion

4.1. Experimental Setup

The system used to train this work in progress complied with the following criteria: AMD Ryzen 7 5800H with a 3.20 GHz Radeon graphics processor. Our system was made up of a 512 GB SSD, a 64-bit operating system, and 32 GB of RAM. An NVIDIA RTX 3050 GPU was used for the investigation.

4.2. Performance Evaluation

Accuracy

The percentage of occurrences in the test dataset that were successfully categorized is gauged here. A class imbalance or a situation in which misclassifying one class would result in more costs than the other would make accuracy deceptive [16]. The Accuracy and loss plots of all competitive models are depicted in Fig.2 and Fig.3, respectively.

\[
\text{Accuracy} = \frac{t_P + t_N}{t_N + t_P + f_N + f_P}
\]  

(1)
Precision and Recall

Precision counts the percentage of real positives among all projected positives, whereas the percentage of actual positives among all observed positives is calculated via recall. When there is a class imbalance, these metrics are helpful since they give a more in-depth view of how each class is doing using the model.

\[
\text{Precision} = \frac{tP}{FP+tP} \quad (2)
\]

\[
\text{Rec} = \frac{tP}{tP+fN} \quad (3)
\]

F1-Score

The F1-score provides a fair representation of the model’s performance for each class and is a harmonic mean of accuracy and recall.

\[
\text{F1 Score} = \frac{2\times(tP)}{2\times tP+FP+fN} \quad (4)
\]

True positives (tP), true negatives (tN), false positives (fP), and false negatives (fN) are the units of measurement for the aforementioned metrics [15].

At different degrees of classification, the trade-off between the genuine positive rate and the false positive rate is shown by the Receiver Operating Characteristic (ROC) curve [14]. It is a useful tool for comparing the performance of many models and choosing the best one. The model’s performance is shown by the Area Under the Curve (AUC) across all categorization thresholds. It stands for the Area under the curve of the ROC curve. The AUC-ROC and Confusion Matrix of several deep learning models are represented in Fig.4 and Fig.5 respectively.

4.3. Comparative Analysis

The effectiveness of numerous pre-trained transfer learning models, including Xception, VGG, ResNet, MobileNet, InceptionV3, and DenseNet models, is emphasized in this section.

In this step, the models were judged using a variety of standards. The total performance of the transfer learning approach is shown in Table 3. VGG19 model outperformed other models in accurately detecting every occurrence of Alzheimer’s disease, with a 93.91% accuracy rate. Contrarily, accuracy rates for VGG16 models and DenseNet201 were 93.59% and 92.34% respectively.

Table 3. Metric measurement of several transfer learning models

<table>
<thead>
<tr>
<th>MODEL</th>
<th>Training Accuracy (%)</th>
<th>Validation Accuracy (%)</th>
<th>Test Accuracy (%)</th>
<th>Precision (%)</th>
<th>Recall (%)</th>
<th>F1 Score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VGG-16</td>
<td>93.59</td>
<td>86.94</td>
<td>80.38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VGG-19</td>
<td>93.87</td>
<td>82.67</td>
<td>76.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ResNet101</td>
<td>79.51</td>
<td>57.45</td>
<td>53.17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MobileNet</td>
<td>70.73</td>
<td>64.71</td>
<td>58.63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>InceptionV3</td>
<td>54.57</td>
<td>47.38</td>
<td>40.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DenseNet201</td>
<td>93.91</td>
<td>93.91</td>
<td>93.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DenseNet121</td>
<td>93.59</td>
<td>93.59</td>
<td>93.59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DenseNet169</td>
<td>92.94</td>
<td>92.94</td>
<td>92.94</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Conclusion and Future Remark

The transfer learning model VGG19 is recommended for the diagnosis of Demented, Non-Demented, Moderately Demented, or Very Mildly Demented brain disorders to attain high classification accuracy [17]. To complete the training and testing process, we require a large enough dataset. The pre-processing methods were used to improve data purification and image scaling. This strategy quickly and favorably impacted all of the models under consideration. The final stage in model building is to expose the CNN transfer learning models [18]. We test the proposed model using the dataset’s 6400 MRI pictures. With a total accuracy of 93.91%, the suggested strategy produced the greatest results for diagnosing Alzheimer’s AD [19].

In the future, to boost the accuracy of the proposed model, we will increase the number of MRI images in the dataset. Future studies might look at more medical image types such as computed tomography (CT), ultrasound, and X-ray using the method outlined here. Additional deep learning techniques can then improve the system’s performance (such as data augmentation and GAN) in future studies [20].
References


