National College of Ireland

Early Detection of Alzheimer's using Deep Learning Technique

MSc Research Project
Data Analytics

Bhagya Vinod
Student ID: 22106111

School of Computing
National College of Ireland

Supervisor: Dr. Christian Horn

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Bhagya Vinod

Student ID: x22106111

Programme: MSc Data Analytics **Year:** 2023

Module: Research in Computing

Supervisor: Dr. Christian Horn

Submission Due Date: 14/12/2023

Project Title: Early Detection of Alzheimer's Detection using Deep Learning Techniques

Word Count: 5534 **Page Count:** 19

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Bhagya Vinod
Date: 14/12/2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Early Detection of Alzheimer's using Deep Learning Techniques

Bhagya Vinod
X22106111

Abstract

The neurocognitive changes associated with Alzheimer's disease (AD) are dynamic, that range from normal cognition to mild cognitive impairment (MCI) and ultimately dementia. Early in the course of the disease, an accurate diagnosis of Alzheimer's disease (AD) is crucial for patient care and will become more so as disease-modifying medications become available. Lately, deep learning algorithms have become the primary tool for medical image processing. These algorithms' effective learning capabilities and capacity to handle challenging issues with relative ease make them appropriate for resolving these image processing issues. With the exponential growth of medical data over the past few years, deep learning has become increasingly useful in the health domain. This work reviews the primary deep learning models that have gained popularity in the past few years and are relevant to medical image analysis. Convolutional neural networks (CNN) and Transfer Learning with ResNet50 can be utilized for image classification in order to take benefit of the changes in preclinical structure for the early detection of AD. Using a custom CNN that has been carefully designed and the potent feature extraction powers of transfer learning with ResNet50, this paper offers an extensive approach to the detection of Alzheimer's disease. The methods and findings reported here offer new opportunities for the field of medical imaging for the diagnosis of neurodegenerative diseases.

1. Introduction

1.1 Background and Significance

Alzheimer's disease is neurodegenerative disease that progresses over time and affects millions of people worldwide. The most prevalent form of dementia is AD, Fig 1 (Amrutesh, et al., 2022). People experiencing Alzheimer's disease will constantly have a deterioration in their memory function. It is associated with various forms of dementia, leading to widespread sickness among the elderly population. A person suffering from dementia will not be able to interact with others or adjust to his surroundings. It is challenging to convey the state of agony they are in. Their ability to process information and remember things will continue to deteriorate and they will depend more and more on others to complete everyday tasks. People

with AD may need care and assistance all the time. Regardless of the high likelihood of AD infection, no proven treatment or cure has been discovered yet. The final observation and the incidence of AD are separated by a considerable amount of time. People who are mildly cognitively impaired (MCI) are those who have early-stage Alzheimer's disease. Most MCI cases result in AD, though only about one-third of those who are infected will develop it.

Prompt and precise diagnosis of Alzheimer's disease is essential for facilitating preventive measures, enhancing the quality of life for those impacted, and maximizing the use of healthcare resources. Proactive treatments are made possible by early Alzheimer's disease diagnosis, which helps patients, and their families better anticipate and manage symptoms. Prompt interventions not only improve quality of life but also lower overall healthcare costs because they typically involve fewer resources in the early stages.

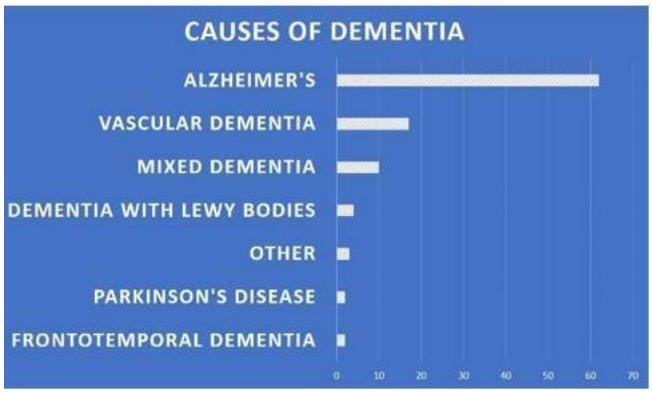

Fig 1. Dementia's causes

The application of deep learning models to medical models to medical image analysis has become a viable method for Alzheimer's disease early detection in recent years. With a focus on integrating Convolutional Neural Networks (CNNs) and Transfer Learning, this project aims to advance the field of Alzheimer's detection by utilizing advanced deep learning techniques. This research paper focuses on the application of Convolution Neural Network and Transfer Learning. Deep learning algorithms have been particularly well-suited to the subtle complexities present in neuroimaging datasets due to their remarkable capacity to extract complex patterns and features from medical images. The most advanced form of image analysis is provided by Convolutional Neural Networks, which are well-suited for recognizing the intricate structures found in medical images because they are specifically designed for pattern recognition in images. The ResNet50 architecture, which was first trained on the ImageNet dataset, is used as a pretrained model. This method makes use of ResNet50's feature extraction capabilities to give Alzheimer's detection a particular advantage.

The rest of the paper will unfold methodology, results, and implications of this research project, showcasing the pivotal role of advanced deep learning in reshaping the landscape of Alzheimer's detection.

1.2 Research Question

How can deep learning approaches be used in order to reliably and rapidly detect Alzheimer's disease?

2. Related Work

2.1 Background

Alzheimer's disease is the most prevalent type of dementia (Liu, et al., 2023),(Singh, et al., 2020), making up 60-80% of all cases (Przedborski, et al., 2003) (Jo, et al., n.d.). It gradually destroys brain cells, hinders thinking and memory and ultimately causes loss of basic cognitive abilities. Over time, the symptoms worsen and develop more slowly (AFZAL, et al., n.d.). The main risk factor for AD is advanced age (beyond 65) (Giorgio, et al., 2020), but it is not solely related to age and affects women more frequently than men. In the United States, over 4.7 million people above the age of 65 have been diagnosed with Alzheimer's disease, according to recent data from the World Alzheimer's Association (Giorgio, et al., 2020). By 2050, this figure is expected to rise to 152 million, with one AD case occurring every three seconds (A., et al., 2020). AD is predicted to cost $1 trillion annually, and estimates suggest that by 2030, this amount will have doubled.

According to (AFZAL, et al., n.d.) stages of Alzheimer can be divided into the following: i) Very Mild Impairment: Memory loss, which is normal as people age. But in some cases, this results in AD. Very mild impairments can cause cognitive problems that interfere with everyday life, such as forgetfulness, uncertainty, personality changes and difficulties with routine tasks. (ii) Mild Impairment: Patients in this stage have more complex daily lives and need more support and care. The symptoms are more severe and resemble very mild impairment. Patients may experience major personality changes, such as becoming irritable or paranoid without cause, and may require assistance with basic tasks like combing their hair. Disorders related to sleep are also common. (iii) Moderate to Severe Impairment: At this point, symptoms could get worse. Individuals may become non-communicative and need round-theclock care. In addition, they might become unconscious and incapable of doing simple tasks like sitting in a chair or keeping their head up. The average time to go from mild to severe Alzheimer's is four years.

Since AD has no known cure (AFZAL, et al., n.d.), early diagnosis is essential for enabling timely treatment, reducing medical costs, enabling accurate diagnosis through the use of modern diagnostic techniques, preventing the disease from getting too severe, while improving the quality of life for the patient (Giorgio, et al., 2020).

2.2 Review of Convolution Neural Network and Transfer Learning with ResNet50

(Dharshini, et al., 2022) highlights the significance of early detection through changes in brain's structure that are visible in preclinical stages. Convolutional, max-pooling, and DenseNet169 layers are used in the CNN architecture to connect layers through dense blocks to effectively extract features. The study investigates the difficulties in acquiring a sizable dataset because of data limitations and privacy issues, and it suggests strategies such as data augmentation. The model's performance over epochs is demonstrated through accuracy and loss graphs that illustrate the results. The suggested strategy offers a viable means of identifying AD early on,

improving public health and wellbeing in the process. (DH & S, 2021) highlights important diseases investigated with Deep Learning and provides an overview across multiple healthcare areas. The study's main objective is to provide a framework for deep learning and predictive analysis-based healthcare data monitoring. (Zaabi, et al., 2020) considers the role of medical imaging methods, specifically Magnetic Resonance Imaging (MRI), in Alzheimer's Disease (AD) diagnosis. This paper focuses on early detection of Alzheimer's disease through deep learning methods for image classification. Alzheimer's disease is a neurodegenerative disease that causes cognitive decline. In the future, the technique could be used to investigate deep quantization as a means of reducing CNN parameters, as well as other brain diseases like cancer. In general, the research offers significant perspectives on utilizing sophisticated imaging and deep learning methods for efficient early identification of Alzheimer's disease. (Abdaelbassat & Hariri, 2023) suggests a method for diagnosing Alzheimer's disease using neuroimaging data that is based on pre-trained CNN models with parameter optimization.

Two key transfer learning approaches are covered in (Geoff, et al., 2023), one involves transferring insights from CNN components to another layer, while the other employs pretrained CNN models for feature extraction. By using an efficient parameter design, the upgraded CNN model reduces the risk of overfitting while increasing computational efficiency and prototype development. High-level and low-level characteristics are derived at once by the model through the division of convolution layers into simultaneous streams. The paper's conclusion highlights the discriminative potential of parameters or features transferred from pre-trained CNNs for classifier training. CNN feature-trained classifiers perform better especially with thoughtful feature transformation and classification. Performance gains are also facilitated by modifying the CNN architecture and enhancing the classifier.

The study (Samanta, et al., 2023) advocates for early detection and categorization of AD phases using neural network algorithms. In order to improve model accuracy and evaluation metrics, the methodology places a strong emphasis on the significance of balanced data, suitable filters, and accurate image preprocessing. The approach motivates researchers and students to investigate more advanced preprocessing methods in order to continuously improve accuracy, especially when used alongside with Transfer Learning. The long-term goal is to create a comprehensive Alzheimer Detection that lets image uploads for continuous patient progress representation. This approach signifies a promising stride in leveraging advanced technologies for the early and accurate detection of Alzheimer's disease.

(Umer, et al., 2023) highlights the increasing significance of deep learning algorithms in the analysis of medical images, with their effectiveness attributes to their ability to learn rapidly and adapt to complicated situations. Deep learning in healthcare is increasing because of the increase in availability of medical data. The study establishes an intense focus on deep learning technology and its functionality. The paper brings out deep learning domination because of its ability to handle data and provide effective solutions.

The study (Devi, et al., 2021) embraces recent advances in deep learning and prioritizes architectural engineering over conventional feature engineering, using the ResNet50 architecture for brain tumour classification. The TL method achieves high accuracy in multiclassifying brain tumours for early diagnosis by dynamically adjusting hyperparameters visa grid search. The effectiveness of Convolutional Neural Networks in image analysis is the main topic of (Rajeshwari, 2021) which also points out that CNNs perform better than conventional machine learning algorithms. The paper (Zaabi, et al., 2020) explores image classification, the study examines the use of CNN and TL for the diagnosis of Alzheimer's Disease (AD), concentrating on two different approaches. It is explained that the convolution layer performs convolutions with the underlying image by moving a filter across the entire image. Transfer Learning is presented as an effective technique for classifying images that makes use of information gathered from a particular architecture to address new challenges.

3. Methodology

This section covers the proposed approach.

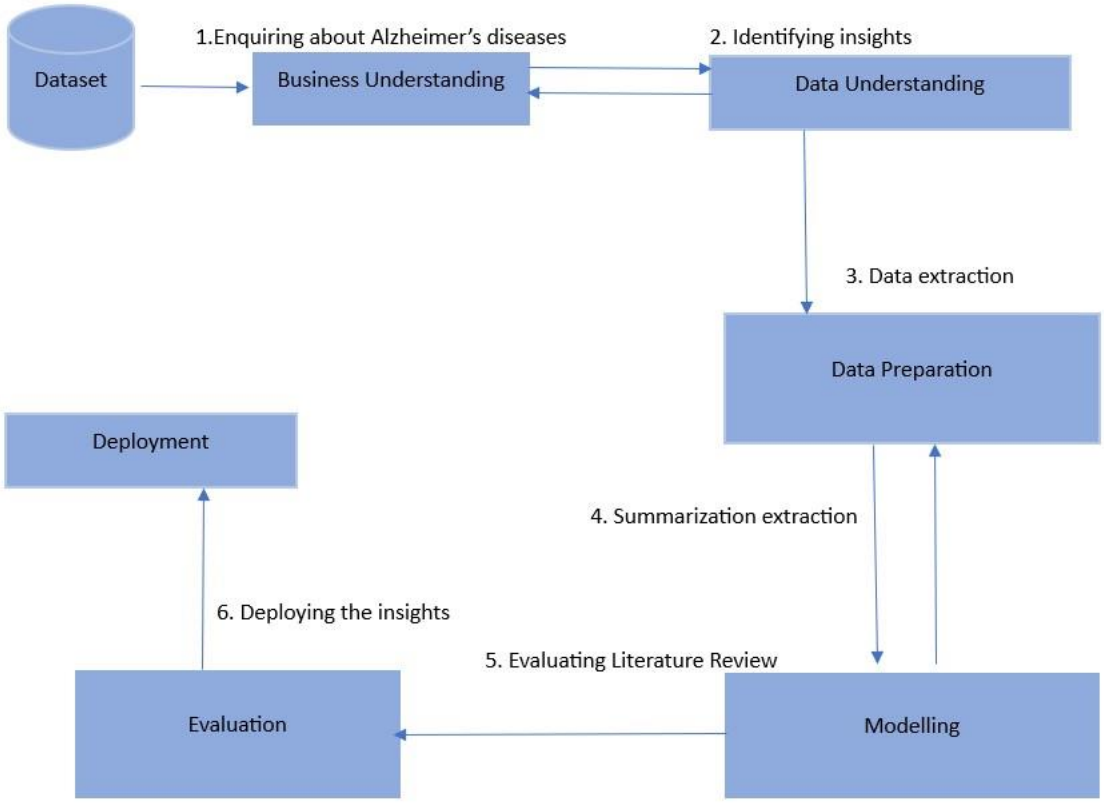

Figure3.1: Methods followed to conduct the study

3.1 Business Understanding

Deep Learning approaches for the early diagnosis of Alzheimer's disease have significant business value as they enable timely intervention, improve quality of life, and reduce overall

healthcare expenses. Proactive treatments made possible by prompt identification help patients and their families more effectively prepare for and control symptoms. As a result, the financial burden on the healthcare system is lessened because early intervention frequently requires fewer resources. Additionally, it aids in the formulation of personalized therapy programs and supports clinical studies and drug development by identifying qualified individuals, enhancing the standard of care given. Early detection also contributes to ongoing research advancements, ethical considerations, public health planning, caregiver support, and research advancements, these outcomes indicate a comprehensive approach that encompasses corporate social responsibility.

3.2 Dataset Description

This dataset is classified into 4 categories, “No impairment, Very Mild Impairment, Mild Impairment and Moderate impairment a mixture of real and synthetic axial MRIs. Each patient’s brain was divided into 32 horizontal axial MRIs for each of the categories which had 100,70,28 and 2 patients respectively. The dataset was sourced from Kaggle.

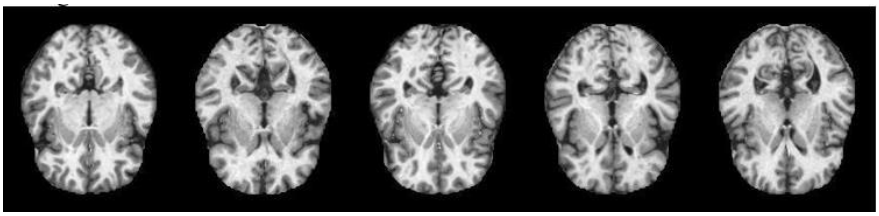

Fig 3.2(1) Sample Images under “No impairment”

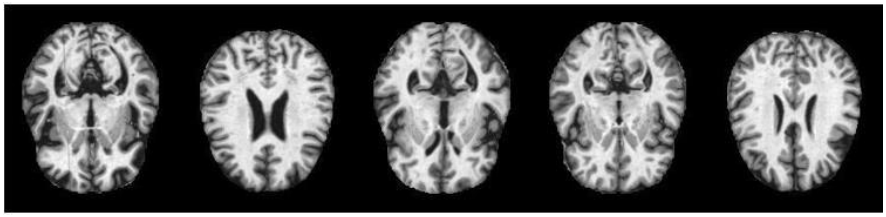

Fig 3.2(2) Sample Images under "Very Mild Impairment”

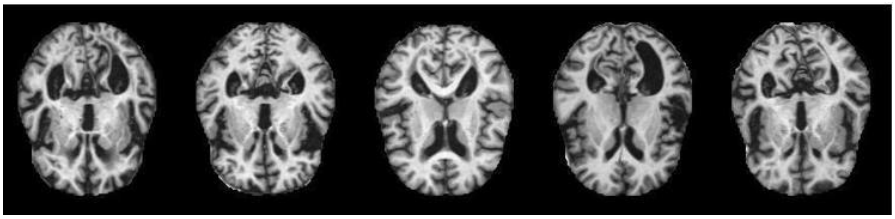

Fig 3.2(3) Sample Images under "Mild Impairment”

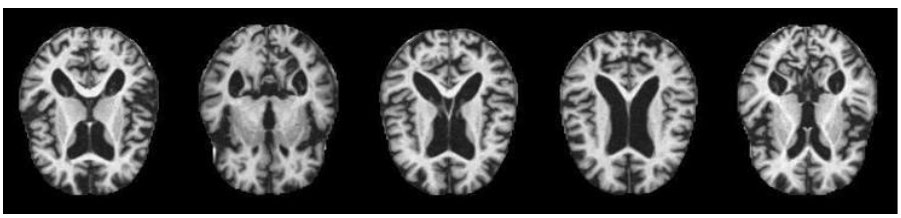

Fig 3.2(4) Sample Images under "Moderate Impairment”

The MRI pictures were obtained using a 1.5 Tesla MRI scanner and a T1-weighted sequence. The images are in ".jpg" format and have a resolution of 128*128 pixels. The skull has been pre-proce[1]ssed out of all photos.

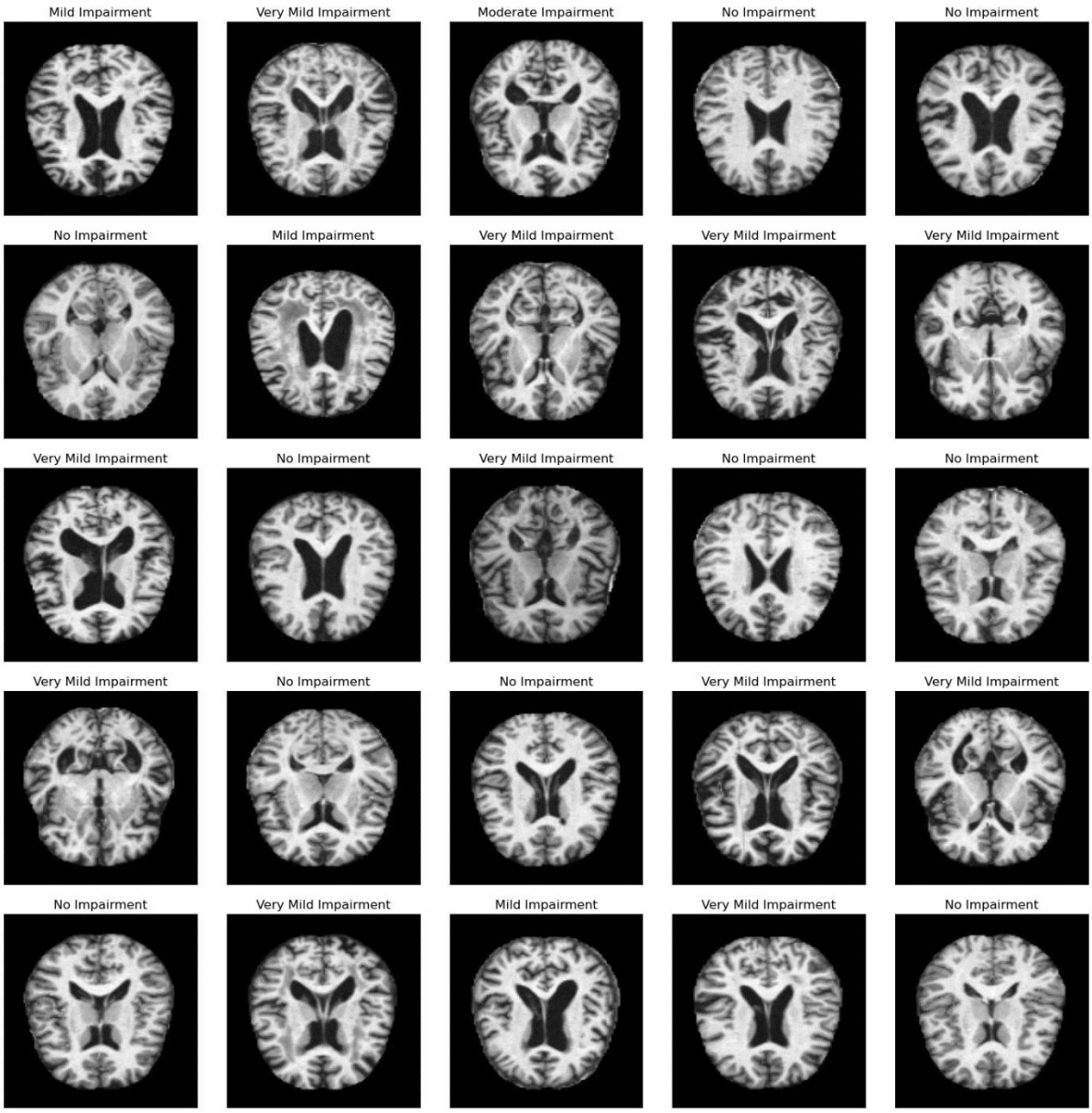

Fig 3.2 (5) Test Image Visualisation

[1] https://www.kaggle.com/datasets/lukechugh/best-alzheimer-mri-dataset-99-accuracy/code

3.3 Data Preparation and Pre-processing

In order to prepare the image dataset for deep learning model training, data preparation and preprocessing are essential. The image files in the dataset have been split into training and testing. In order to ensure a random distribution throughout training, the data is shuffled. Preprocessing involves loading the images and applying rescaling techniques to normalize the outcome to a range of 0 to 1. TensorFlow's ImageDataGenerator class is utilized for real-time data augmentation, thereby minimizing overfitting, and expanding the dataset. These preparatory steps make sure the data is properly formatted, augmented, and well-organized in order to enhance the model's generalization abilities.

Pre-processing images in the train and test sets independently was accomplished by creating "train_generator" and "test_generator" objects using Keras's Image Data Generator class. All input images have the same dimensions 128*128 pixels. Neural networks require fixed-sized input images, so this was significant. Since many activation functions are made to function between 0 and 1, the parameter "rescale" was set to (0,1) to prevent outbursts in gradients and to improve performance. In order to interpret the grayscale MRIs as three channel RGB images, the parameter "color_mode" was set to "rgb", as the majority of deep learning models have been designed to work with RGB images, especially those that are meant for image analysis and classification.This was crucial because RGB images are required as input for pretrained models. In order to combine the class labels into one hot encoded vector- a prerequisite for the softmax activation function used in multiclass classification, the parameter "class_mode" was set to "categorical". Images were converted into batches of 32 image arrays when the parameter "batch_size" was set to 32. This was significant for CNNs as it enables regularization of the training process, parallel computing with GPUs and effective processing of large datasets. To guarantee a fair evaluation and comparison, the parameter "shuffle" was set to "False". This ensures that the images in the train and test sets are arranged in the exact same order each time a multiple model is trained or evaluated.

3.4 Exploratory Data Analysis

EDA is an essential stage in the Alzheimer detection project that helps to comprehend and get the dataset ready for efficient model development. The below histogram represents the count instances belonging to each class. The x-axis displays the class labels, and each bar has an annotation with the corresponding counts above it. This visualization offers significant insight into potential challenges and factors to consider during model evaluation and performance assessment. It also helps to understand the balance or imbalance of class distribution within the testing dataset. The histogram displays the precise number for every class, giving detailed insights into the distribution of the dataset.

Fig 3.4(1) displays the number of samples for several classes, from which we can see that the most common class in the testing dataset is 'No Impairment' with 640 samples, followed by 'Very Mild Impairment' with 448 samples. The 'Mild Impairment' class has 179 samples, and the 'Moderate Impairment' class has the fewest samples, with only 12.

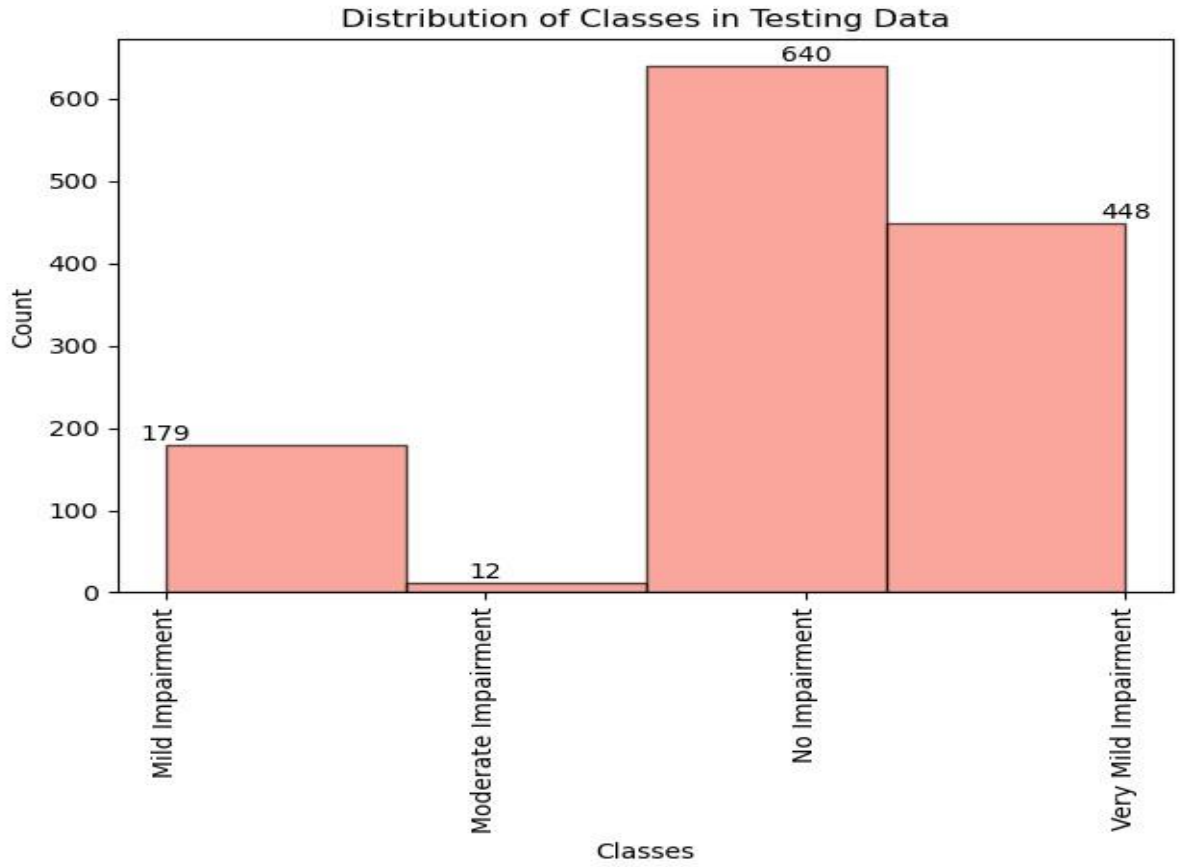

Fig 3.4(1) Distribution of images belonging to each class

3.5 Model Selection

In this project, model selection contributes a critical aspect that influences accuracy and performance of the classification task. A custom convolutional neural network (CNN) architecture is used in this implementation because it can capture complex patterns and features in medical imaging data and greater flexibility. A custom network is used here for clarity and to highlight the fact that the architecture is not a pre-built solution. In order to improve generalization, enhance feature extraction and mitigate overfitting, the model architecture is designed with a series of convolutional layers, batch normalization, and dropout layers. In light of the special qualities of medical imaging datasets, where minute patterns and structures can be essential for a precise diagnosis, this custom architecture was selected. The model is enhanced with dense layers for effective classification and makes use of softmax activation function in the output layer for multi-class classification. The choice of this model demonstrates a careful assessment of the needs and characteristics of the dataset at hand and is consistent with the project's focus on maximising efficiency in the context of Alzheimer's disease diagnosis.

4. Design Specification

The design specification demonstrates a methodical and intentional approach to developing a convolutional neural network and transfer learning using ResNet50 for the purpose of detecting Alzheimer's disease. In order to emphasize the reproducibility of results, parameters such as random seeds, epochs, batch size are defined. File paths and labels have been carefully extracted from the file system and arranged into a pandas DataFrame, demonstrating the dataset's meticulous structure. Multiple convolution layers with different filter sizes, activation functions and batch normalization are part of the CNN model architecture, which improves feature extraction. Dropout layers are placed to avoid overfitting, the final dense layers culminate in a softmax activation for multi-class classification. Efficient loading and processing of images are guaranteed using image data generators from TensorFlow's Keras API. A set of evaluation metrics, including accuracy, AUC, and F1 score are combined with the Adam optimizer and categorical crossentropy loss to compile the model. Early Stopping and model checkpointing are incorporated into the training process to maximize performance and prevent overfitting. In order to provide a reliable and efficient method for Alzheimer's disease detection, this design specification emphasizes the meticulous consideration of parameters, data handling, model architecture, and evaluation metrics.

The Transfer Learning model is constructed using the ResNet50 architecture, pretrained on the dataset, applying its feature extraction capabilities. The base ResNet50 model is initialized with weights from the pre-training phase, and its top layers are excluded to allow for the addition of custom dense layers suited for the classification task. The Sequential API from Keras is used to define the model's subsequent layers. The relevant information is effectively summarized by reducing the spatial dimensions of the feature maps using a Global Average Pooling 2D Layer. Rectified Linear Unit (ReLU) activation functions are employed to activate each of the dense layers, that have 1024 and 512 units. After each dense layer, dropout layers with a rate of 0.2 are added to minimize overfitting by arbitrarily eliminating a portion of the input units during training. The softmax function activates the final dense layer, which consists of four units, which represent the number of classes in the classification task, and outputs class probabilities.

5. Implementation

5.1 Data Preparation

The image data is organized into training and testing datasets by reading the file paths of images stored in specified directories and creating corresponding dataframes.

5.2 Custom CNN Model Building

A custom Convolutional Neural Network is built using TensorFlow and Keras. This model includes several layers**, Conv2D Layers** for feature extraction from images, **MaxPooling2D Layers** for downsampling the feature maps, **BatchNormalization Layers** to normalize the activations and gradients, **Dropout Layers** for regularization to prevent overfitting, **Dense Layers** fully connected layers for classification.

As multiclass classification was being done, all the pretrained models and the custom CNN were assembled using Adam as the optimizer, with a learning rate of 0.001, “Accuracy” and “F1 Score” as the metrics that will be calculated after each epoch, and “Categorical Cross Entropy” as the loss function. Early stopping with patience of 10 and model checkpoint were used for callbacks, and both were designed to track validation loss. Since patience should be 10% of the total epochs, early stopping was used to prevent the models from overfitting, and 10 was chosen as the value. A model checkpoint was added so that each model would be loaded with weights for which its validation loss was the lowest at the end of training. Finally, custom CNN and all of the pretrained models were trained for 100 epochs by validating their performance on the test data after each epoch and updating their weights using the training data. For image classification, the model architecture is organized in a sequential manner and its layers are systematically configured. The input layer is modified to support images with three color channels (RGB) and a size specified by the variable IMAGE_SIZE. Subsequent layers comprise of convolutional blocks, max-pooling for spatial downsampling, batch normalization for stable training and dropout for regularization. In order to speed up convergence, batch normalization is added throughout the convolutional layers, that utilize rectified linear unit (ReLU) activation functions. The model emphasizes hierarchical feature extraction by gradually reducing spatial dimensions through max-pooling. A softmax activation in the output layer, signifying a multiclass classification task with four classes, concludes the final layers, which are densely connected with varying numbers of neurons and dropout rates. The large number of parameters in the model emphasizes its complexity and highlights its ability to learn complex patterns. The model can effectively mitigate overfitting and generalize to unseen data due to batch normalization and dropout layers. The ImageDataGenerator is a powerful tool for efficiently loading and preprocessing images during the training of neural networks. Train_generator and test_generator are two image data generators created. For training and testing purposes, these generators will be used to preprocess and load batches of images, respectively. The data generators have successfully found and processed the images, 10,240 images in the training set and 1,279 images in the testing set. The Adam optimizer with a learning rate of 0.0001 is utilized to optimize the model during training. Categorical accuracy, area under the ROC curve and F1 score are the chosen metrics for evaluation.

Model: "cnn_model"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 128, 128, 16)	448
conv2d_1 (Conv2D)	(None, 128, 128, 16)	2320
max_pooling2d (MaxPooling2D)	(None, 64, 64, 16)	0
conv2d_2 (Conv2D)	(None, 64, 64, 32)	4640
conv2d_3 (Conv2D)	(None, 64, 64, 32)	9248
batch_normalization (BatchNormalization)	(None, 64, 64, 32)	128
max_pooling2d_1 (MaxPooling2D)	(None, 32, 32, 32)	0
conv2d_4 (Conv2D)	(None, 32, 32, 64)	18496
conv2d_5 (Conv2D)	(None, 32, 32, 64)	36928
batch_normalization_1 (BatchNormalization)	(None, 32, 32, 64)	256
max_pooling2d_2 (MaxPooling2D)	(None, 16, 16, 64)	0
conv2d_6 (Conv2D)	(None, 16, 16, 128)	73856
conv2d_7 (Conv2D)	(None, 16, 16, 128)	147584
batch_normalization_2 (BatchNormalization)	(None, 16, 16, 128)	512
max_pooling2d_3 (MaxPooling2D)	(None, 8, 8, 128)	0
conv2d_8 (Conv2D)	(None, 8, 8, 256)	295168
last_conv_layer (Conv2D)	(None, 8, 8, 256)	590080
batch_normalization_3 (BatchNormalization)	(None, 8, 8, 256)	1024
max_pooling2d_4 (MaxPooling2D)	(None, 4, 4, 256)	0
flatten (Flatten)	(None, 4096)	0
dropout (Dropout)	(None, 4096)	0
dense (Dense)	(None, 512)	2097664
batch_normalization_4 (BatchNormalization)	(None, 512)	2048
dropout_1 (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 128)	65664
batch_normalization_5 (BatchNormalization)	(None, 128)	512
dropout_2 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 64)	8256
batch_normalization_6 (BatchNormalization)	(None, 64)	256
dropout_3 (Dropout)	(None, 64)	0
dense_3 (Dense)	(None, 4)	260

Total params: 3355348 (12.80 MB)
Trainable params: 3352980 (12.79 MB)
Non-trainable params: 2368 (9.25 KB)

Fig 5.1 (1) Custom CNN architecture

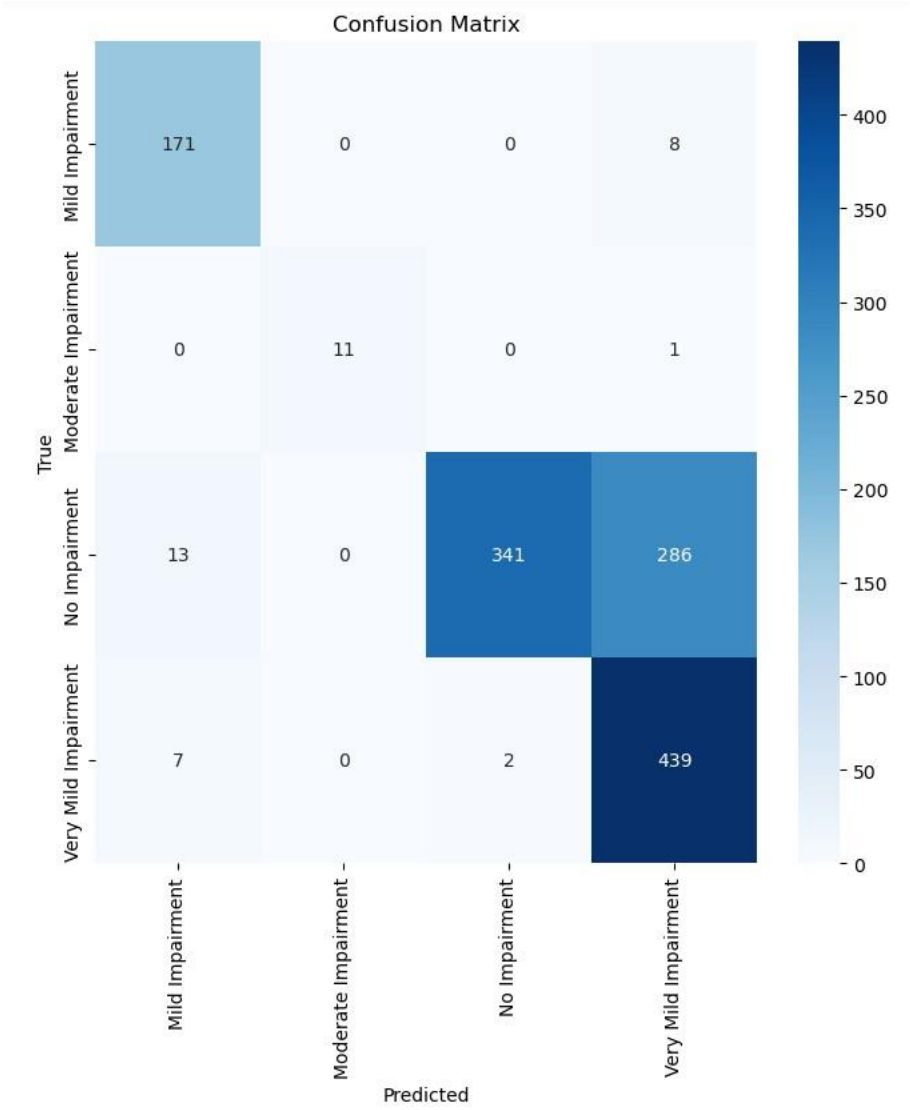

Fig 5.1 (2) Confusion Matrix of CNN Model

5.3 Transfer Learning with ResNet50

The pre-trained ResNet50 model is used as the base model with its layers set to non-trainable. Additional layers are added on top:

- **GlobalAveragePooling2d Layer**: To reduce the feature map size.
- **Dense Layers**: For classification
- **Dropout Layers**: To reduce overfitting

As a feature extractor, the pre-trained ResNet50 model uses the ImageNet dataset and is configured to exclude its top classification layer. After that, this base model is frozen to guarantee that the features it has learned will not be altered in a subsequent training session. The Sequential model, named 'transfer_learning_model' is constructed by adding the pretrained ResNet50 as the base layer, followed by a global average pooling layer to reduce spatial dimension. A final softmax layer with four units for classification is added, along with two densely connected layers for regularization that have ReLU activation and dropout. The

model is compiled with the Adam optimizer with a learning rate of 0.001, categorical crossentropy as the loss function, and additional metrics such as categorical accuracy, AUC, and F1 score for evaluation. The training is performed on a set of images with specified epochs and callbacks for monitoring. Once training is completed, the model is evaluated on the test set and testing accuracy is calculated.

6. Evaluation

The primary stage when developing a deep learning model is to evaluate its performance. To evaluate the performance, standard metrics such as accuracy, precision, recall and F1 score that play a crucial role in assessing the model's ability to correctly identify instances of Alzheimer's disease are employed. The evaluation process involved partitioning the dataset into testing and training sets, ensuring robust generalization. The results obtained from the evaluation contribute valuable insights into the reliability and applicability of CNN and Transfer Learning with ResNet50 in the early detection of Alzheimer's disease, paving the way for future advancements in the field.

6.1 Experiment with CNN

The CNN output details the training process of a Convolutional Neural Network (CNN) over several epochs for a binary classification task. A complete set of performance metrics such as loss, accuracy (acc), Area Under the Curve (AUC) and F1 score are provided for each epoch for the training and validation datasets. As indicated in fig 6.1(1), training ceases after 27 epochs as the validation loss has not improved for the prior 10 epochs, indicating that an early stopping mechanism is in place to prevent overfitting. The model's performance is summarized in the final epoch, which indicates the best validation loss that was achieved and the corresponding accuracy, AUC, and F1 score.

```
Epoch 24: val_loss did not improve from 0.16999
320/320 [==============================] - 175s 547ms/step - loss: 0.0329 - acc: 0.9899 - auc: 0.9994 - f1_score: 0.9899 - v
al_loss: 3.6875 - val_acc: 0.4464 - val_auc: 0.6629 - val_f1_score: 0.3120
Epoch 25/100
320/320 [==============================] - ETA: 0s - loss: 0.0472 - acc: 0.9846 - auc: 0.9989 - f1_score: 0.9846
Epoch 25: val_loss did not improve from 0.16999
320/320 [==============================] - 173s 541ms/step - loss: 0.0472 - acc: 0.9846 - auc: 0.9989 - f1_score: 0.9846 - v
al_loss: 1.3466 - val_acc: 0.7123 - val_auc: 0.8953 - val_f1_score: 0.6277
Epoch 26/100
320/320 [==============================] - ETA: 0s - loss: 0.0406 - acc: 0.9881 - auc: 0.9992 - f1_score: 0.9881
Epoch 26: val_loss did not improve from 0.16999
320/320 [==============================] - 173s 542ms/step - loss: 0.0406 - acc: 0.9881 - auc: 0.9992 - f1_score: 0.9881 - v
al_loss: 0.8556 - val_acc: 0.7936 - val_auc: 0.9380 - val_f1_score: 0.7727
Epoch 27/100
320/320 [==============================] - ETA: 0s - loss: 0.0534 - acc: 0.9818 - auc: 0.9989 - f1_score: 0.9818
Epoch 27: val_loss did not improve from 0.16999
320/320 [==============================] - 175s 547ms/step - loss: 0.0534 - acc: 0.9818 - auc: 0.9989 - f1_score: 0.9818 - v
al_loss: 0.9830 - val_acc: 0.7522 - val_auc: 0.9173 - val_f1_score: 0.8294
Epoch 27: early stopping
```

Fig 6.1(1) training process and early stopping of CNN model

The results of the model evaluation demonstrates and equal trade off, with a loss of 0.930, an accuracy of 75.22%, an Area Under the Curve (AUC) of 0.9173, and an F1 score of 0.8294. The model's ability to classify instances in the test dataset is demonstrated by its testing accuracy of 75.22%. The below fig 6.1(2) provides an in-depth analysis, including precision, recall, F1-score and support metrics for each of the four classes: "Mild Impairment", "Moderate

Impairment", "No Impairment"," Very Mild Impairment". The proportion of positive identifications that were accurate are represented by precision, while proportion of actual positive correctly identified are indicated by recall. The F1-score, being the harmonic mean of precision and recall, provides a balanced measure of a model's performance. The support metrics represent the number of actual occurrences of each class in the dataset. Additionally, macro, and weighted averages are provided, offering a comprehensive assessment across all classes.

The "Mild Impairment" category has a high precision of 90% indicating that when the model predicts Mild Impairment. Recall implies that suggests that 96% of the real cases of mild impairment have been identified by the model. The balanced performance is further supported by the F1-Score 92%. Collectively, these metrics demonstrate that the model can effectively classify Mild Impairment while minimizing false positive, with support from 179 instances.

The "Moderate Impairment" group has very good performance indicators. With a precision of 1.00, every prediction made for this class was accurate. With a recall of 0.92, the model was able to correctly identify 92% of the real cases of moderate impairment. It is further highlighted by the high F1-score 0.96 that recall and precision should be balanced. These metrics show the model's strong ability to classify Moderate Impairment accurately while maintaining high precision, with a support of 12 instances.

The "No Impairment" class exhibits high precision (0.99), meaning that instances classified as No Impairment were accurately predicted. On the contrary hand, the recall (0.53) is comparatively low, indicating that a sizable fraction of true No Impairment cases were missed by the model. The trade-off is highlighted by the F1-Score (0.69), which displays a balance between recall and precision. These metrics, implemented with the support of 640 instances, demonstrate the model's accuracy in predicting No Impairment instances, but they also point to areas where recall could be improved.

The model's moderate precision of 0.60 for the "Very Mild Impairment" class suggests that it can reasonably predict cases classified as such. The high recall (0.98) suggests that most real cases of Very Mild Impairment are well captured by the model. The F1-Score (0.74) indicates a good overall performance in this class by establishing a balance between recall and precision. Collectively, these metrics demonstrate the model's efficiency in predicting cases of Very Mild Impairment, with a support of 448 instances.

```
40/40 [==============================] - 8s 193ms/step - loss: 0.9830 - acc: 0.7522 - auc: 0.9173 - f1_score: 0.8294
Testing Accuracy: 75.22%
40/40 [==============================] - 8s 187ms/step
                    precision    recall  f1-score   support

     Mild Impairment       0.90      0.96      0.92       179
 Moderate Impairment       1.00      0.92      0.96        12
       No Impairment       0.99      0.53      0.69       640
Very Mild Impairment       0.60      0.98      0.74       448

            accuracy                           0.75      1279
           macro avg       0.87      0.85      0.83      1279
        weighted avg       0.84      0.75      0.75      1279
```

Fig 6.1(2) Classification Report

6.2 Experiment with Transfer Learning using ResNet50

The training progress is displayed for each epoch during the training phase. For each epoch, the output includes the values of various metrics, such as loss, accuracy (acc), area under the ROC curve(auc), and F1 score. These metrics are essential for monitoring the performance of the model during training. The training was stopped early after 30 epochs to prevent overfitting. The validation metrics, including validation loss, accuracy, auc, and F1 score are also displayed in fig. The Transfer Learning Model Testing Accuracy is reported as 51.52%, indicating the proportion of correctly classified instances in the test dataset.

```
Epoch 28/100
320/320 [==============================] - ETA: 0s - loss: 0.7399 - acc: 0.6643 - auc: 0.8995 - f1_score: 0.6563
Epoch 28: val_loss did not improve from 0.16999
320/320 [==============================] - 321s 1s/step - loss: 0.7399 - acc: 0.6643 - auc: 0.8995 - f1_score: 0.6563 - val_
loss: 0.9402 - val_acc: 0.5262 - val_auc: 0.8253 - val_f1_score: 0.4442
Epoch 29/100
320/320 [==============================] - ETA: 0s - loss: 0.7263 - acc: 0.6690 - auc: 0.9034 - f1_score: 0.6614
Epoch 29: val_loss did not improve from 0.16999
320/320 [==============================] - 318s 993ms/step - loss: 0.7263 - acc: 0.6690 - auc: 0.9034 - f1_score: 0.6614 - v
al_loss: 0.9103 - val_acc: 0.5512 - val_auc: 0.8346 - val_f1_score: 0.4616
Epoch 30/100
320/320 [==============================] - ETA: 0s - loss: 0.7283 - acc: 0.6651 - auc: 0.9016 - f1_score: 0.6563
Epoch 30: val_loss did not improve from 0.16999
320/320 [==============================] - 335s 1s/step - loss: 0.7283 - acc: 0.6651 - auc: 0.9016 - f1_score: 0.6563 - val_
loss: 0.9488 - val_acc: 0.5152 - val_auc: 0.8200 - val_f1_score: 0.4008
Epoch 30: early stopping
40/40 [==============================] - 42s 1s/step - loss: 0.9488 - acc: 0.5152 - auc: 0.8200 - f1_score: 0.4008
Transfer Learning Model Testing Accuracy: 51.52%
40/40 [==============================] - 44s 1s/step
```

Fig 6.2 (1) training process of Transfer Learning

The classification report fig 6.2(2) provides a detailed breakdown of the model's performance for each class in the test dataset. It includes precision, recall, F1-score, and support for each class, "Mild Impairment", "Moderate Impairment", "No Impairment"," Very Mild Impairment".

For the class "Mild Impairment", the model exhibited a precision of 0.34, indicating that 34% of the instances predicted as Mild Impairment were indeed true positives. The recall, measuring the proportion of actual Mild Impairment instances correctly predicted by the model, was 0.55. The F1- Score, an integrated mean of precision and recall, stood at 0.42. These metrics were calculated based on a support of 179 instances in the Mild Impairment class.

For "Moderate Impairment" category, the model's precision was 0.16 signifying that only 16 % of the cases that were predicted to be in this category actually were. With a recall of 0.75, the model's accuracy in identifying real cases of Moderate Impairment was significantly higher. The harmonic mean of recall and precision, or F1-score was 0.26. Although precision is low, the higher recall indicates that the model was more successful in capturing a sizable portion of the actual cases of Moderate Impairment.

For the category "No Impairment" with a precision of 0.63, the model demonstrates that 63% of the instances that were predicted to have no impairment were in fact true. Recall, a measure of the percentage of real No Impairment cases that the model accurately detected was 0.75. The harmonic mean of recall and precision, or F1-score was 0.68. These metrics were computed based on a dataset support of 640 instances in the No Impairment class. The precision and recall

values are balanced, indicating that the performance in accurately classifying instances that belong to the No Impairment category was reasonable.

The model's precision for the "Very Mild Impairment" category was 0.42 implicating that 42% of the cases that were predicted to fall into this category were accurate. The recall, which measures the percentage of real cases of "Very Mild Impairment" that the model correctly identified was 0.17. The harmonic mean of recall and precision, or F1-score was 0.24. The lower recall value suggests a challenge in correctly identifying a significant portion of the actual Very Mild Impairment instances, which is reflected in the lower F1-score.

	precision	recall	f1-score	support
Mild Impairment	0.34	0.55	0.42	179
Moderate Impairment	0.16	0.75	0.26	12
No Impairment	0.63	0.75	0.68	640
Very Mild Impairment	0.42	0.17	0.24	448
accuracy			0.52	1279
macro avg	0.39	0.55	0.40	1279
weighted avg	0.51	0.52	0.49	1279

Transfer Learning Balanced Accuracy Score: 55.34 %
Transfer Learning Matthew's Correlation Coefficient: 23.51 %

Fig 6.2 (2) Classification Report

6.3 Discussion

CNN Model is more effective for Alzheimer's disease detection, based on the metrics provided. It indicates a more accurate and balanced performance across various classes, which is crucial for medical diagnostics. It is a more dependable option due to its high recall and precision in crucial categories as well as a good balance of these metrics as demonstrated by the F1-score. The CNN model outperforms the Transfer Learning model in almost all metrics. It has higher precision, F1-score, accuracy, balanced accuracy, and MCC. Based on its higher balanced accuracy score, the CNN model seems to be able to better handle class imbalance. The CNN model's higher MCC indicates that its predictions are more likely to come true. The CNN model seems to be more appropriate for Alzheimer's detection, where precise and uniform classification are essential.

Model Name	Balanced Accuracy	Matthews Correlation Coefficient
CNN	84.6%	66.93%
ResNet50	55.34%	23.51%

Table 6.3 (1) Result comparison between the two models

7. Conclusion and Future Work

In the evaluation of Alzheimer's detection models, CNN model outperformed the Transfer Learning model on several metrics. In terms of precision, recall, f1-score, overall accuracy, balanced accuracy, and Matthew's Correlation Coefficient (MCC), the CNN model performs better. The effectiveness of CNN model in Alzheimer's detection is highlighted by its emphasis

on precision and recall in crucial categories as well as its balanced performance as demonstrated by f1-scores. The results imply that the CNN model is the ideal option for medical diagnostics where accurate and consistent classification is essential.

As we conclude our current study on the use of advanced machine learning algorithms for Alzheimer's detection, it is essential to map out a path for future research projects that will expand our understanding, improve our techniques, and promote the ongoing development of diagnostic tools in this essential field. Consider how longitudinal data may be utilized to track changes in brain images over time. Develop models that can evaluate and determine how Alzheimer's disease develops over time, offering insights into the disease's early stages. Look into external variables that affect the accuracy of Alzheimer's detection models, such as lifestyle, socioeconomic status or multiple medical conditions. To improve the predictability of results, consider creating models that take these outside factors into consideration. Perform comprehensive clinical validation investigations to evaluate model's feasible performance across different clinical contexts. Review the opportunities and difficulties that come with using machine learning models in clinical settings.

Acknowledgement

I would like to express my sincere gratitude to my supervisor, Dr. Christian Horn for guiding and encouraging me throughout this experiment. He guided me through each phase of this project, making it achievable.

References

A., H., Badawy, H. M. & Haikal, A. Y., 2020. Deep Learning Approcah for Early Detection of Alzheimer. pp. 2-3.

Abdaelbassat, B. M. & Hariri, W., 2023. Alzheimer Disease Detection using Advanced Transfer Learning Techniques on MRI Images. *International Conference on Decision Aid Sciences and Applications,* pp. 490-493.

AFZAL, S. et al., n.d. A Data Augmentation-Based Framework to Handle Class Imbalance Problem for Alzheimer's Stage Detection.

Amrutesh, A. et al., 2022. Alzheimer Disease Prediction using Machine Learning and Transfer Learning Models. *6th International Conference on Computation System and Information Technology for Sustainable Solutions,* p. 1.

D. D. et al., 2023. Classification and Prediction of Alzheimer's Disease using Deep Learning. pp. 6-9.

Devi, O., Bindu, D. C. S. & Kumar, D. E. S., 2021. *IDENTIFICATION AND CLASSIFICATION OF BRAIN TUMOR FROM MRI USING TRANSFER LEARNING APPROACH.* s.l.:s.n.

Dharshini, S. S., Mallika, R. N. & Peesapati, S. N., 2022. Deep Learning based Detection Model for Alzheimer's Disease. pp. 6-9.

DH, C. & S, V. S., 2021. Alzheimer's Disease Detection from Brain MRI Data using Deep Learning Techniques. *2nd Global Conference for Advancement in Technology,* pp. 2-8.

Giorgio, J. et al., 2020. Modelling prognostic trajectories of cognitive decline due to Alzheimer's disease. pp. 6-9.

Jo, T., Nho, K. & Saykin, A. J., n.d. Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction using Neuroimaging Data.

Liu, S. et al., 2023. Multimodal Neuroimaging Feature Learning for Multiclass Diagnosis of Alzheimer’s Disease. *IEEE Transactions on Biomedical Engineering,* pp. 1132- 1134.

Pershiya, S. et al., 2023. Alzheimer's Disease Diagnosis using Deep Learning Approach. pp. 4-6.

Przedborski, S., Vila, M. & Lewis, V. J.-., 2003. Neurodegeneration: What is it and where are we. pp. 3-9.

Rajeshwari, S. S., 2021. ATransfer Learning Approach dor Predicting Alzheimer's Disease. *International Conference on Nascent Technologies in Engineering.*

Samanta, S., Mazumder, I. & Roy, C., 2023. Deep Learning based Early Detection of Alzheimer's Disease using Image Enhancement Filters. *hird International Conference on Advances in Electrical.*

Singh, S. P. et al., 2020. 3D Deep Learning on Medical Images: A Review. pp. 3-9.

Umer, M., Sharma, S. & Rattan, D. P., 2023. *A Survey of Deep Learning Models for Medical Image Analysis.* s.l.:s.n.

Wang, T. & Cao, L., 2021. Deep Learning Based Diagnosis of Alzheimer's Disease Using Structural Magnetic Resonance Imaging: A Survey. *International Conference on Applied Machine Learning,* pp. 408-411.

Zaabi, M., Smaoui, N., Derbel, H. & Hariri, W., 2020. Alzheimer's disease detection using convolutional neural networks and transfer learning based methods. pp. 2-4.