Alzheimer’s Disease Classification Using MRI Images: A Deep Learning Approach



Sami Ul Haq

2020-KIU-BS2025

**&**

Danish Ali

2020-KIU-BS2006

**Department of Computer Science, Faculty of Natural Science and Engineering, KIU, Gilgit.**

**September, 2024**

Alzheimer’s Disease Classification Using MRI Images: A Deep Learning Approach

By

Sami Ul Haq

2020-KIU-BS2025

**&**

Danish Ali

2020-KIU-BS2006

A thesis submitted in partial fulfillment of the requirement for the degree of BS

in

Computer Sciences

Department of Computer Science, Faculty of Natural Science and Engineering, KIU, Gilgit.

September, 2024



In the Name of Allah Almighty

The Most Beneficent

The Most Merciful

# CERTIFICATE OF APPROVAL

This Project “Alzheimer’s Disease Classification Using MRI Images: A Deep Learning Approach” is hereby approved in partial fulfillment of the requirement for the degree of Bachelor of Science in Computer Sciences, Karakoram International University, Gilgit-Baltistan.

Thesis Supervisor ­­­­­­­­­­­­\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Dr. Dostdar Hussain)

Chairperson/Chairman/HOD \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Dr. Zarnawab Khan)

External Examiner 1 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

External Examiner 2 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# DEDICATION

This thesis is dedicated to my parents, without whose unconditional love, sacrifices, and unwavering support I would not have achieved the achievements I have. Thank you very much for your constant encouragement, belief in my aspirations, and all of the sacrifices that enabled me to achieve his goal. This thesis is also dedicated to my friends, who have always inspired and supported me, as well as to all of the teachers who have supplied me with a dominant orientation in life. Thank you for helping, motivating, and guiding me on this long path.

# DECLARATION

We hereby declare that this thesis/project is a presentation of our own work and that it has not been submitted anywhere for any award. We also warrant, that we have not received outside assistance or involved the external contributions, if received/involved we will acknowledge in written statement to authorities, otherwise we will be liable for the cancellation of our thesis thereby the degree that will be awarded.

|  |  |
| --- | --- |
| Scholar Name: SAMI UL HAQ  Reg: 2020-KIU-BS2025  Date: 20-SEP-2024  Signature | Supervisor Name: DR. DOSTDAR HUSSAIN  Date: 20-SEP-2024  Signature |
| Scholar Name: DANISH ALI  Reg: 2020-KIU-BS2006  Date: 20-SEP-2024  Signature |  |

# ACKNOWLEDGMENTS

First and foremost, I am eternally grateful to Allah Almighty for endowing me with the patience, intellect, and ability to comprehend and complete this study project. I'd want to take this occasion to offer my sincere gratitude to my supervisor. **Dr. Dostdar Hussain, Assistant Professor in deportment of Computer Science** whose encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of the subject and completion of this research.

I would also like to extend gratitude to my friends and specially seniors for their help and motivation throughout my thesis work.

Sami Ul Haq

Danish Ali

# ABSTRACT

Alzheimer's disease is a degenerative condition in which the gradual impairment of cognitive functions and mental abilities extends over an extended period, thus, determining its stages is crucial for appropriate care and treatment. This study proposal attempts to identify the stages of Alzheimer's disease through MRI scans based on applications of advanced deep learning methodologies like DenseNet and ResNet, and the dataset used for such research was spread across the identified four types. These groups consist of patients with Very Mild Dementia, Mild Dementia, and Moderate Dementia. All these require data augmentation and oversampling in order to obtain a balanced distribution for analytical purposes. The researchers tested the model using measurements of accuracy, precision levels, as well as recall and F-score measurements. Besides these issues related to the class distribution, the findings obtained from the research of testing on the basis of the used dataset indicate that deep learning-based models can classify the overall stages of Alzheimer's disease. It shows a bright future towards the improvement in the diagnosis procedure.

Alzheimer’s Disease Classification Using MRI Images: A Deep Learning Approach (2024)

# Table of Contents

Chapter Page

[CERTIFICATE OF APPROVAL 4](#_Toc177686373)

[DEDICATION 5](#_Toc177686374)

[DECLARATION 6](#_Toc177686375)

[ACKNOWLEDGMENTS 7](#_Toc177686376)

[ABSTRACT 8](#_Toc177686377)

[Table of Contents 10](#_Toc177686378)

[LIST OF FIGURES 12](#_Toc177686379)

[LIST OF TABLES 13](#_Toc177686380)

[ACRONYMS AND ABBRIVATION 14](#_Toc177686381)

[Chapter 1 15](#_Toc177686382)

[INTRODUCTION 15](#_Toc177686383)

[1.1. Background 15](#_Toc177686384)

[1.2. Project Overview 16](#_Toc177686385)

[1.3. Scope and Limitations 16](#_Toc177686386)

[1.4. Statement of the Problem 17](#_Toc177686387)

[Chapter 2 19](#_Toc177686388)

[**REVIEW OF LITERATURE** 19](#_Toc177686389)

[2.1. Introduction 19](#_Toc177686391)

[2.1.1. Deep Learning in Medical Imaging 19](#_Toc177686392)

[2.1.2. Related Work on Alzheimer’s Classification 20](#_Toc177686393)

[2.1.3. Analysis of CNN and VGG19 Models 20](#_Toc177686394)

[2.2. Conclusion 25](#_Toc177686395)

[Chapter 3 26](#_Toc177686396)

[**Methodology** 26](#_Toc177686397)

[3.1. Data Collection 26](#_Toc177686399)

[3.2. Data Preprocessing 26](#_Toc177686400)

[3.3. Model Selection 28](#_Toc177686401)

[3.4. Model Training and Validation 29](#_Toc177686402)

[3.5. Evaluation Metrics 29](#_Toc177686403)

[Chapter 4 30](#_Toc177686404)

[**System Architecture and Model Implementation** 30](#_Toc177686405)

[4.1. Hardware and Software Specifications 30](#_Toc177686407)

[4.2. Model Architecture Details 31](#_Toc177686408)

[4.2.1. ResNet50 31](#_Toc177686409)

[4.2.2. DenseNet121 32](#_Toc177686410)

[4.3. Training Configuration and Hyperparameter Tuning 33](#_Toc177686411)

[Chapter 5 35](#_Toc177686412)

[**Results and Analysis** 35](#_Toc177686413)

[5.1. ResNet50 35](#_Toc177686415)

[5.1.1. AUC and Loss Curves of ResNet50 35](#_Toc177686416)

[5.1.2. Confusion Matrix ResNet50 35](#_Toc177686417)

[5.1.3. Classification Report of ResNet50 36](#_Toc177686418)

[5.2. DenseNet121 37](#_Toc177686419)

[5.2.1. AUC and Loss Curves of DenseNet121 37](#_Toc177686420)

[5.3. Comparison of Models (ResNet50 vs. DenseNet121) 40](#_Toc177686421)

[5.4. Discussion on Overfitting Issues 40](#_Toc177686422)

[Chapter 6 41](#_Toc177686423)

[**Conclusion and Future Work** 41](#_Toc177686424)

[6.1. Conclusion 41](#_Toc177686431)

[6.2. Future Work 41](#_Toc177686432)

[REFERENCES 42](#_Toc177686433)

[Appendix A 43](#_Toc177686434)

[Appendix B 45](#_Toc177686435)

[Appendix C 47](#_Toc177686436)

# LIST OF FIGURES

Figure 1 - VGG19 AUC CURVES 21

Figure 2 – VGG19 LOSS CURVES 21

Figure 3 – VGG19 CONFUSION MATRIX 22

Figure 4 - CNN AUC CURVES 24

Figure 5 - CNN LOSS CURVES 24

Figure 6 - CNN CONFUSION MATRIX 25

Figure 7 - Number of Images in each class 26

Figure 8 - No of Images After Augmentation 28

Figure 9 - ResNet50 Model AUC and Loss Curves 35

Figure 10 - ResNet50 Confusion Matrix 36

Figure 11 - DensNet Model AUC and Loss Curves 38

Figure 12 - DensNet Confusion Matrix 39

# LIST OF TABLES

[**Table 1 - VGG19 CLASSIFICATION REPORT** 24](#_Toc177651414)

[**Table 2 - RESNET50 CLASSIFICATION REPORT** 37](#_Toc177651415)

[**Table 3 - DENSNET CLASSIFICATION REPORT** 40](#_Toc177651416)

# ACRONYMS AND ABBRIVATION

KIU KARAKORAM INTERNATIONAL UNIVERSITY

CS COMPUTER SCIENCES

AD Alzheimer’s disease

MRI Magnetic Resonance Imaging

CNN Convolution Neural Network

# Chapter 1

## INTRODUCTION

## Background

Alzheimer's disease is a neurodegenerative progressive disorder that primarily manifests with cognitive decline, characterized by loss of memory and an impaired thinking ability, which is predominant among older adults. It is one of the most common forms of dementia and accounts for 60-80% of cases worldwide. Early detection and classification of Alzheimer's disease would be crucial in proper intervention and management, especially because the disease goes into several stages: MCI, early-stage AD, and advanced-stage AD. This needs to be told promptly and correctly so that results may be achieved, and quality of life for patients must be improved.

Recent developments in the field of medical imaging, especially concerning Magnetic Resonance Imaging (MRI), have yielded significant understanding of the morphological alterations in the brain linked to Alzheimer's disease. Nevertheless, conventional diagnostic methods frequently exhibit subjectivity and depend on the interpretation by specialists, resulting in inconsistencies in diagnosis. In response to these issues, deep learning algorithms have surfaced as effective instruments for the automated classification of Alzheimer's disease. For instance, CNNs like ResNet50 and DenseNet121 are very efficient at analyzing neuroimaging data. This is because these can find hierarchical patterns from highly complex image datasets.

Recent studies indicate that deep learning algorithms are capable of attaining significant accuracy in differentiating among the various stages of Alzheimer's disease. ResNet50 and DenseNet121 are the models used the most, because they have a high capability of feature extraction, and superior results have been shown by the models in AD detection from MRI scans. Deep residual learning framework in ResNet50 helps to reduce the vanishing gradient issue and it is apt for further extensive features extracted from the MRI images. On the other side, DenseNet121 connects each layer with all the subsequent layers that enhance the flow and reuse of feature maps but might not be precise in some instances while compared to ResNet50.

There is a need for objective, scalable, and efficient diagnostic tools that induce the integration of deep learning into the diagnosis of AD.

Accuracy also rose up to a level of almost 87% for models like ResNet50, and the DenseNet121 models reached an accuracy rate around 75%. This concludes the kind of ability they hold but simultaneously calls for improvement. Current studies are in the phase of fine-tuning these models, optimizing the hyperparameters, and exploring hybrid approaches combining the CNN with traditional machine learning techniques, such as SVM, to enhance their diagnostic accuracy along with dependability. And through the automation of the classification process, these deep learning models reduce the clinician workload, thereby promoting early and accurate identification of Alzheimer's disease towards the initiation of treatment as well as proper management of a patient. Such improvement is on its way to revolutionize the neuroimaging diagnosis area of Alzheimer's into a new hope for both patients and professionals.

## Project Overview

This project focuses on the classification of Alzheimer's disease stages using deep learning models to improve diagnostic accuracy. Alzheimer's disease is characterized by various stages: Non Demented, Very Mild Demented, Mild Demented, and Moderate Demented, with the dataset consisting of 2197, 2151, 861, and 62 images, respectively. Due to imbalanced class distribution, previous works have reported accuracies below 70%. To address these challenges, this project employed data augmentation and oversampling techniques to balance the dataset. The models used, including ResNet50 and DenseNet121, achieved improved accuracies of 87% and 75%, respectively. This work aims to leverage advanced deep learning techniques to enhance the early detection and classification of Alzheimer's disease, providing a significant step towards better clinical outcomes.

## Scope and Limitations

**Scope:**  
The scope of this project includes the use of advanced deep learning models like ResNet50 and DenseNet121 to classify Alzheimer's disease stages using MRI images. The study focuses on enhancing the diagnostic accuracy by addressing class imbalance through data augmentation and oversampling techniques. These methods were applied to a four-class dataset containing Non-Demented, Very-Mild-Demented, Mild-Demented, and Moderate-Demented images. The primary objective is to improve model performance, aiding in early detection and stage classification of Alzheimer’s, which is crucial for timely intervention and treatment planning.

**Limitations:**  
Despite the improved overall performance, the project faced significant challenges due to the severe imbalance in the dataset, particularly in the Mild and Moderate Demented classes, which have a very low number of images. While data augmentation and oversampling helped to balance the classes synthetically, they cannot fully replicate the diversity and complexity of real-world data. As a result, the models still struggle with accurately predicting these underrepresented classes. The reliance on synthetic data augmentation methods highlights a critical limitation: the need for more comprehensive, real-world datasets to achieve reliable performance across all Alzheimer’s stages. Additionally, the models used are computationally intensive, requiring significant processing power, which may limit their applicability in real-time clinical settings. Future work should focus on collecting larger, balanced datasets and exploring lightweight model architectures to overcome these constraints.

## Statement of the Problem

Alzheimer's disease (AD) is a progressive neurological disorder with varying stages of cognitive decline, making early and accurate diagnosis crucial for effective management. However, current diagnostic methods, including traditional imaging analysis, often struggle to accurately differentiate between the various stages of AD, especially in cases with subtle structural changes. This project addresses the problem of classifying Alzheimer's stages using deep learning models on an imbalanced MRI dataset with four classes: Non Demented, Very Mild Demented, Mild Demented, and Moderate Demented.

A significant challenge lies in the severe class imbalance, particularly in the Mild and Moderate Demented categories, which contain significantly fewer images compared to the Non Demented and Very Mild Demented classes. This imbalance leads to poor model performance in predicting the underrepresented classes, hindering the accuracy and reliability of stage classification. While data augmentation and oversampling techniques were employed to synthetically balance the dataset, these methods do not fully capture the complexity and variability of real-world data, resulting in limited improvements.

Therefore, the core problem this project addresses is the need for a robust deep learning approach capable of handling imbalanced datasets and improving stage-specific classification performance. The overarching goal is to enhance the diagnostic accuracy of Alzheimer's disease classification models, thereby contributing to better clinical decision-making and patient outcomes. However, to achieve truly reliable results, there is an urgent need for larger, balanced, and high-quality datasets that reflect the diversity of the disease across all stages.

# Chapter 2

**REVIEW OF LITERATURE**



## Introduction

Alzheimer’s disease (AD) [1] is a progressive neurodegenerative disorder characterized by cognitive decline, memory loss, and impaired reasoning, which severely affect daily functioning. It is the most common form of dementia, accounting for 60-80% of cases worldwide. AD primarily affects older adults, but early-onset cases can occur. Pathologically, it is marked by abnormal protein accumulations, including amyloid-beta plaques and tau tangles, leading to neuronal damage and brain atrophy.

The disease progresses through various stages: Non Demented, Very Mild Demented, Mild Demented, and Moderate Demented. Early and accurate detection of these stages is essential for timely intervention and improved quality of life. However, conventional diagnostic approaches, such as clinical assessments and neuroimaging, often face challenges in differentiating between these stages due to subtle and overlapping symptoms. This has created a critical need for automated, precise classification methods to support clinical decisions.

Recent advancements in artificial intelligence (AI) and deep learning [2] offer promising approaches for enhancing diagnostic accuracy through the analysis of MRI scans. These techniques can potentially identify patterns that are often missed by human evaluation, facilitating the classification of AD stages. The ongoing research focuses on leveraging deep learning models, such as CNNs and VGG architectures, to improve classification performance and aid in the early detection of Alzheimer's, which is crucial for better management and care strategies. Despite these advancements, achieving reliable results remains a challenge, especially when working with imbalanced datasets and limited data in specific classes.

## Deep Learning in Medical Imaging

Deep learning has transformed medical imaging [3], particularly with Convolutional Neural Networks (CNNs), which excel in analyzing complex visual data for disease detection and classification. In Alzheimer’s research, deep learning models like CNNs, ResNet, and DenseNet analyze MRI scans to detect subtle brain changes associated with disease stages, outperforming traditional methods. Despite its promise, challenges such as data imbalance, overfitting, and interpretability issues persist. Research continues to refine these models to improve accuracy and integrate them effectively into clinical practice.

## Related Work on Alzheimer’s Classification

Recent studies have focused on deep learning models like CNN [4] and VGG19 [5] for Alzheimer's classification using MRI data. Existing models have achieved moderate success, with reported accuracies often below 70%, largely due to challenges such as data imbalance and limited sample sizes in key dementia stages. For instance, one study utilizing CNN achieved an accuracy of 63%, while another employing VGG19 reached 66%. These findings highlight the need for more sophisticated models and balanced datasets to improve classification performance and support clinical applications effectively.

## Analysis of CNN and VGG19 Models

In this section, we compare the performance of the Convolutional Neural Network (CNN) and VGG19 models for Alzheimer's disease classification, focusing on their accuracy, loss, and confusion matrix to evaluate their effectiveness.

1. **VGG19 Model Analysis**

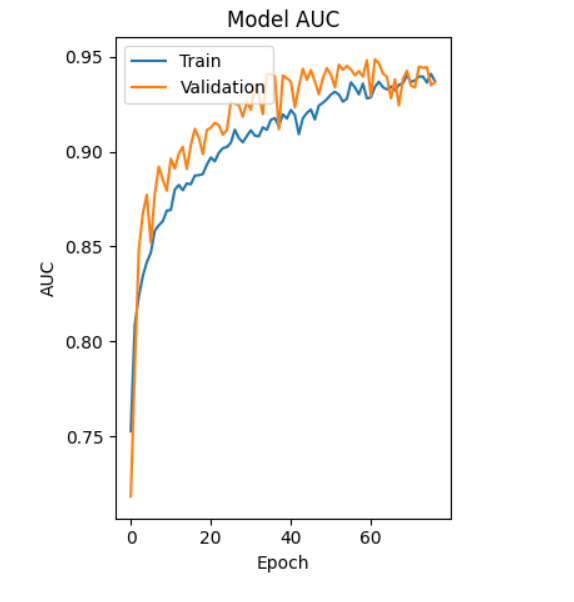
The VGG19 model, a deeper and more complex architecture, was employed to further explore the classification of Alzheimer’s disease. Although it demonstrated high training performance, the model exhibited signs of overfitting, as evidenced by discrepancies between training and testing accuracies.

* + 1. **Training and Validation Accuracy**

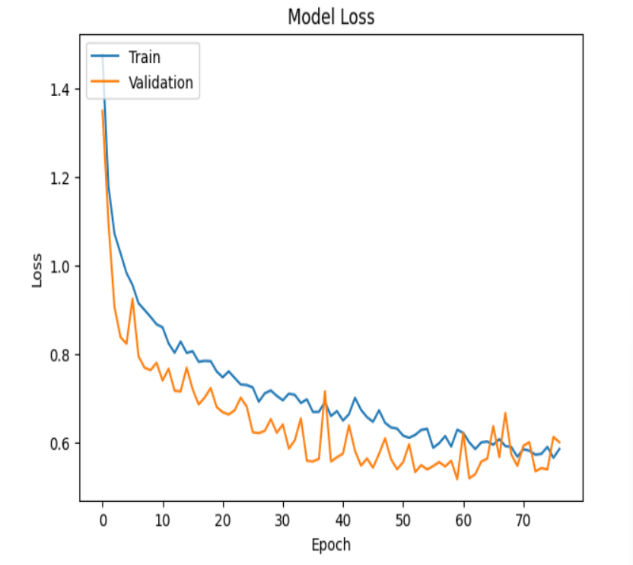
The VGG19 accuracy plot (**Figure 1**) shows that both training and validation accuracies reach above 90% early in the training process. However, despite these high accuracies, the model's generalization capability is limited, with the testing accuracy dropping to 66%, indicating that the model has overfitted to the training data.

* + 1. **Training and Validation Loss**

The loss curve for VGG19 (**Figure 2**) aligns with the accuracy analysis, where training loss continually decreases, but the validation loss fluctuates, suggesting unstable learning. The validation loss does not converge as smoothly, reinforcing the overfitting issue.



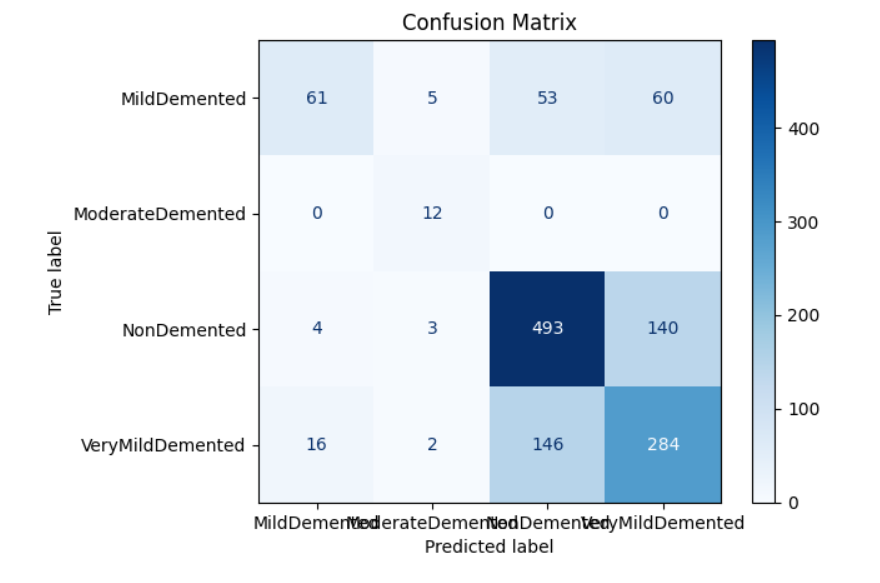
**Figure 1 - VGG19 AUC CURVES**



**Figure 2 – VGG19 LOSS CURVES**

* + 1. **Confusion Matrix Analysis**

The confusion matrix (**Figure 3**) further highlights the model’s difficulties in generalizing across classes, especially in distinguishing between Mild and Very Mild Demented categories. Misclassifications are prevalent, particularly in classes with fewer samples, emphasizing the impact of class imbalance on model performance.

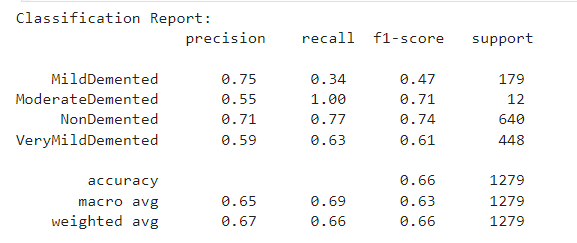


**Figure 3 – VGG19 CONFUSION MATRIX**

**Figure-3**

* + 1. **Classification Report**

The detailed classification report shows varying performance across the four categories, with Non Demented and Very Mild Demented achieving relatively higher precision and recall compared to the Mild and Moderate Demented classes. The weighted average accuracy of 66% underscores the need for further adjustments.

**Table 1 - VGG19 CLASSIFICATION REPORT**

1. **CNN Model Analysis**

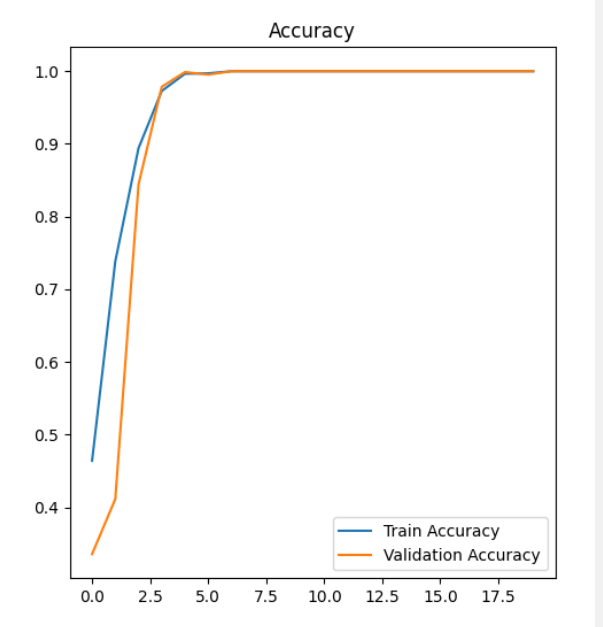
The CNN model exhibits significant overfitting, with training accuracy reaching nearly 100% while testing accuracy stagnates at 66%. This discrepancy suggests that the model memorizes training data but struggles to generalize to unseen data.

**Training and Validation Accuracy**

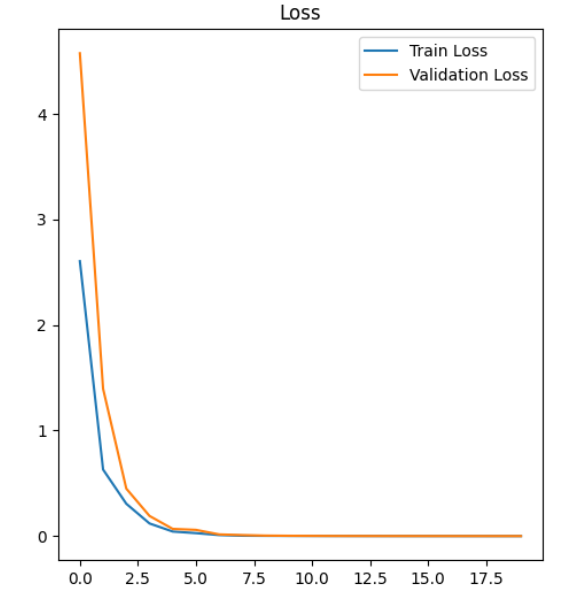
The accuracy plot (**Figure 4**) shows an extremely high training accuracy, which quickly approaches 100%, while the validation accuracy remains low at around 66%, illustrating overfitting.

**Training and Validation Loss**

The loss graph (**Figure 5**) shows training loss approaching zero, indicating overfitting, while the validation loss fluctuates, reflecting poor model generalization.



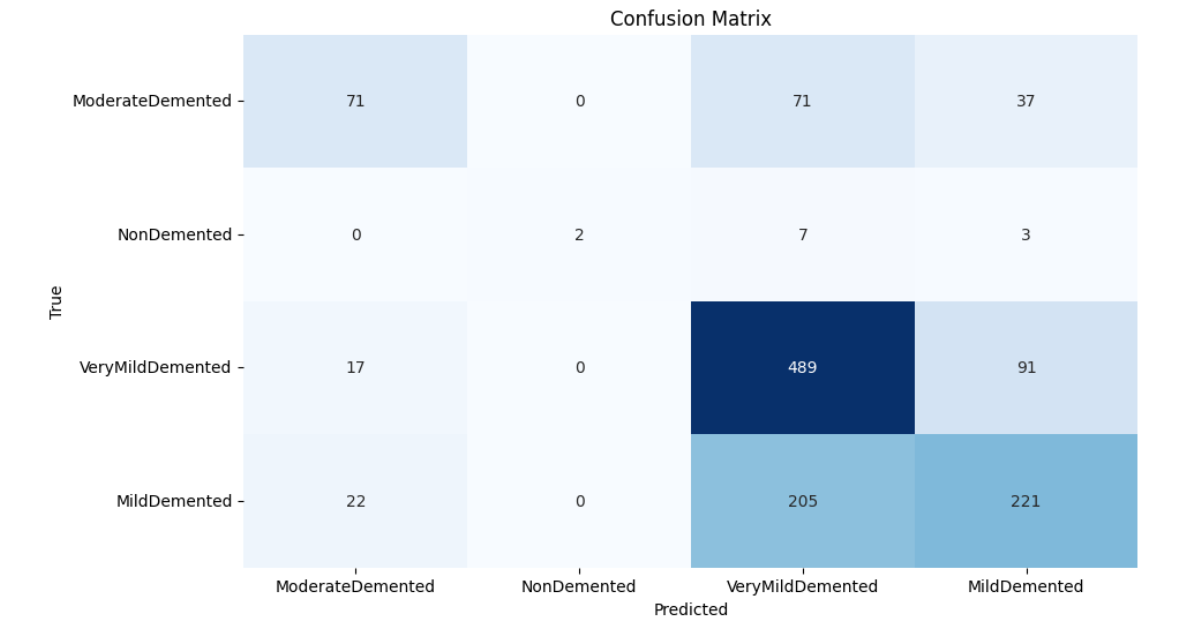
**Figure 4 - CNN AUC CURVES**



**Figure 5 - CNN LOSS CURVES**

**Confusion Matrix Analysis**

The confusion matrix (**Figure 6**) confirms that the CNN model misclassifies a significant number of instances, reflecting an inability to generalize beyond the training data.



**Figure 6 - CNN CONFUSION MATRIX**

## Conclusion

The reviewed literature highlights several challenges in Alzheimer’s classification, primarily due to imbalanced datasets that often skew model performance. The previously used models, such as CNN and VGG19, struggled with overfitting and low generalizability, particularly when faced with complex and imbalanced data distributions. These models exhibited high training accuracy but poor testing accuracy, reflecting their limitations in capturing the nuances of underrepresented classes. Addressing these issues, such as employing advanced data augmentation techniques or class balancing strategies, could significantly enhance the performance and robustness of future models in Alzheimer's disease classification.

# Chapter 3

**Methodology**



## Data Collection

The dataset utilized in this study was sourced from Kaggle, a well-known platform for publicly available datasets. It contains MRI images [6] categorized into four classes representing different stages of Alzheimer’s: Non Demented (2,197 images), Very Mild Demented (2,151 images), Mild Demented (861 images), and Moderate Demented (62 images). The significant imbalance in class distribution poses a challenge, as deep learning models tend to perform poorly on underrepresented classes. Therefore, addressing this imbalance was a primary focus during data preprocessing to ensure robust model training and generalization.

**Figure 7 - Number of Images in each class**

## Data Preprocessing

Data preprocessing was a crucial step that involved several tasks to enhance the dataset's quality [4] and address class imbalance, ensuring effective training of the deep learning models.

**Image Resizing and Normalization**

Initially, all images were resized to 224x224 pixels to match the input requirements of the deep learning models (DenseNet121 and ResNet50). Each pixel value was normalized to fall within the range [0, 1], facilitating faster convergence and improving model performance.

**Data Augmentation and Balancing**

To address the class imbalance and increase the number of training samples, data augmentation [7] techniques were employed. These included:

* **Rotation**: Randomly rotating images by varying degrees.
* **Flipping**: Applying horizontal and vertical flips to create mirrored versions of images.
* **Zooming and Shifting**: Applying zoom and translation to simulate different perspectives.

These techniques helped generate synthetic variations of existing images, significantly enhancing the dataset's diversity and providing the models with more generalized features. Specifically, the data augmentation process increased each class to 3,000 images, which included balancing the Mild and Moderate Demented classes that were originally underrepresented.

**Oversampling**

Oversampling was applied specifically to the Mild Demented and Moderate Demented classes to match the number of images in other classes. This step was crucial to avoid training bias toward the majority classes and ensured that the models would not underperform on the less frequent conditions.

**Figure 8 - No of Images After Augmentation**

**Data Splitting Strategy**

After balancing the dataset through augmentation and oversampling, the data was split into training and testing sets using an 80-20 split.

**Training Set**: Each class was split to have 2,400 images in the training set, used to train the models.

**Testing Set**: Each class had 600 images allocated to the testing set, which was used to evaluate the model's performance.

This split ensured that the training set was diverse enough for the models to learn meaningful patterns, while the test set provided an unbiased evaluation of the models' ability to generalize to unseen data. This strategic data handling played a critical role in enhancing model performance, especially on minority classes.

## Model Selection

For this study, two advanced deep learning models, ResNet and DenseNet, were selected due to their superior performance in image classification tasks, especially in medical imaging. ResNet (Residual Network) is known for its deep architecture and skip connections that prevent vanishing gradient problems, making it particularly effective for extracting complex patterns from medical images. DenseNet (Densely Connected Convolutional Network) emphasizes feature reuse, with each layer receiving inputs from all previous layers, leading to improved efficiency and reduced parameter requirements. Both models were tested to assess their capabilities in accurately classifying the different stages of Alzheimer’s.

## Model Training and Validation

The training phase involved dividing the dataset into training and validation sets to evaluate the models' performance. The models were trained using a range of hyperparameters, including learning rate, batch size, and the number of epochs, which were fine-tuned through a systematic approach to achieve optimal performance. Regularization techniques such as dropout were used to prevent overfitting, and early stopping was implemented to halt training when the validation loss plateaued or worsened. Checkpoints were used to save the best model weights, ensuring that the final model retained the highest accuracy during validation.

## Evaluation Metrics

The performance of the models was evaluated using several metrics, including accuracy, precision, recall, F1-score, and confusion matrices. These metrics provided comprehensive insights into the models' ability to distinguish between the four classes of Alzheimer’s. Precision and recall were particularly important in this context, as they offered a deeper understanding of the model’s effectiveness in identifying the minority classes without being biased towards the more prevalent classes. The confusion matrix further illustrated the classification errors, highlighting areas where the model misclassified one stage as another, thus guiding further improvements in the modeling approach.

# Chapter 4

**System Architecture and Model Implementation**



## Hardware and Software Specifications

Initially, the model was run on a local machine using Jupyter Notebook, but due to the computationally intensive nature of deep learning models, this setup proved inefficient. For instance, training on just five epochs took over 20 hours to complete due to the high parameter count and large dataset size. To overcome this, the experimentation moved to Google Colab and Kaggle, which offer GPU acceleration (e.g., NVIDIA T4), significantly reducing training time and improving efficiency.

**Hardware Specifications:**

* **Local Machine:**

CPU-based, not efficient for deep learning tasks due to slow processing speeds and limited memory capacity.

* **Google Colab and Kaggle:**

NVIDIA T4 GPUs were used for their computational power, optimized memory handling, and faster processing of large models with extensive image datasets.

**Software Specifications:**

Python libraries utilized include:

**NumPy (v1.26.4):** For numerical computations and array manipulation.

**Pandas (v2.2.2):** For data manipulation and analysis.

**TensorFlow (v2.15.0):** Core framework for deep learning model development.

**Matplotlib (v3.7.5):** For plotting graphs and visualizing data.

**PIL (Python Imaging Library):** For image processing tasks.

**Scikit-learn:** Tools like classification\_report and confusion matrix for model evaluation.

These libraries were selected for their latest updates, ensuring compatibility with the latest deep learning frameworks and performance optimizations. The latest versions were confirmed and utilized to enhance execution speed, model compatibility, and access to advanced features.

## Model Architecture Details

## ResNet50

The model architecture is based on the ResNet50 framework, a popular deep learning model pre-trained on the ImageNet dataset. The architecture was chosen due to its depth, skip connections, and proven performance in image classification tasks.

**Base Model (ResNet50):**

The pre-trained ResNet50 was used as the base model. It was loaded without its top layers to allow the addition of custom layers for the specific task of Alzheimer's classification. The base model was frozen to retain the learned weights and prevent retraining.

**Custom Layers:**

* **GlobalAveragePooling2D:** This layer was added to reduce the spatial dimensions of the feature maps from the base model.
* **Dense Layers:** Two fully connected layers were added with 1024 and 512 neurons, both using ReLU activation functions. These layers serve as the classifier and help in learning complex patterns specific to the dataset.
* **Output Layer:** The final layer consists of 4 neurons, each representing a class in the dataset (Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented), with a softmax activation function to output class probabilities.

**Compilation:**

* **Optimizer:** The Adam optimizer was used with a very low learning rate (0.000001) to fine-tune the model gradually, minimizing the risk of overshooting the minima.
* **Loss Function:** Categorical cross-entropy was chosen as the loss function, suitable for multi-class classification problems.
* **Metrics:** Accuracy was used as the primary metric to evaluate model performance during training and testing.

This architecture aims to leverage the feature extraction power of ResNet50 while allowing for the fine-tuning of the model on the specific task of Alzheimer's disease classification.

## DenseNet121

The model architecture is based on the DenseNet121 framework, a deep learning model known for its dense connections between layers and efficient feature reuse, pre-trained on the ImageNet dataset. The architecture was chosen for its ability to capture intricate patterns in images, making it suitable for Alzheimer’s disease classification.

**Base Model (DenseNet121):**

The pre-trained DenseNet121 model was used as the base model. It was loaded without its top layers to allow the addition of custom layers for this specific task. To retain the feature extraction capabilities of DenseNet121, all but the last 20 layers were frozen, preventing them from being retrained during the fine-tuning process.

**Custom Layers:**

* **GlobalAveragePooling2D:** This layer was added to reduce the spatial dimensions of the feature maps produced by the DenseNet121 base model. It effectively reduces the data to a 1024-dimensional feature vector.
* **Dense Layers:**
  1. A dense layer with 1024 neurons was added, initialized with 'he\_uniform' to maintain efficient gradient flow. Batch Normalization was applied to stabilize and speed up the training process, followed by a ReLU activation function for non-linearity. Dropout was introduced to prevent overfitting by randomly setting a fraction of input units to zero during training.
  2. Another dense layer with 512 neurons followed the same setup, further refining the learned features.
* **Output Layer:** The final dense layer consists of 4 neurons, each corresponding to a class in the dataset (Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented). A softmax activation function was used to output class probabilities.

**Compilation**:

* **Optimizer:** The Adam optimizer was employed with a small learning rate, balancing the speed of convergence with the stability of training. The learning rate can be adjusted based on experimental results.
* **Loss Function:** Categorical cross-entropy was chosen as the loss function, as it is well-suited for multi-class classification problems.
* **Metrics:** Accuracy was selected as the primary metric for evaluating model performance during both training and testing phases.

## Training Configuration and Hyperparameter Tuning

**Training Configuration:**

The training was conducted using TensorFlow with two state-of-the-art deep learning architectures: DenseNet121 and ResNet50, each adapted for Alzheimer’s classification. Both models utilized image datasets with four classes: NonDemented, VeryMildDemented, MildDemented, and ModerateDemented. Data augmentation techniques were applied to enhance the dataset’s variability, and datasets were split into training, validation, and test sets to evaluate model performance.

* **Batch Size:** The models were trained with a batch size of 16 due to computational limitations and to improve convergence rates.
* **Image Size:** All input images were resized to 224x224 pixels, which aligns with the input requirements of both DenseNet121 and ResNet50 architectures.
* **Data Augmentation:** Used techniques such as rotation, flipping, and zooming during data preparation to simulate new data points and help models generalize better.

**Hyperparameter Tuning:**

* **Learning Rate:** For both models, an initial low learning rate was chosen. DenseNet121 used 0.00001, while ResNet50 used 0.000001, ensuring gradual adjustments in weights, stabilizing training, and preventing large, disruptive updates.
* **Freezing Layers:** For DenseNet121, the initial layers were frozen, except the last 20 layers, allowing only these to be trainable. In ResNet50, all base model layers were frozen to leverage pre-trained knowledge from ImageNet while training the added dense layers.
* **Class Weights:** Calculated using training dataset class distributions to handle class imbalance and ensure the models do not bias towards more prevalent classes.
* **Optimizer:** Both models used the Adam optimizer, known for adaptive learning rates and handling sparse gradients efficiently.
* **Metrics and Loss Functions:** The models were trained using categorical cross-entropy as the loss function, suitable for multi-class classification tasks, and monitored with metrics like accuracy and AUC to assess predictive performance.
* **Callbacks:**
  1. **Model Checkpoint:** Monitors validation metrics and saves the best model weights to avoid overfitting and maintain the highest accuracy.
  2. **Early Stopping:** Terminates training when validation performance ceases to improve, optimizing training time and preventing overfitting.
  3. **Learning Rate Scheduler (DenseNet121):** Reduces learning rate upon plateauing validation loss, ensuring steady progress even when learning slows.

Overall, these configurations and tuning strategies were essential for achieving optimal performance in classifying Alzheimer’s stages with the chosen models.

# Chapter 5

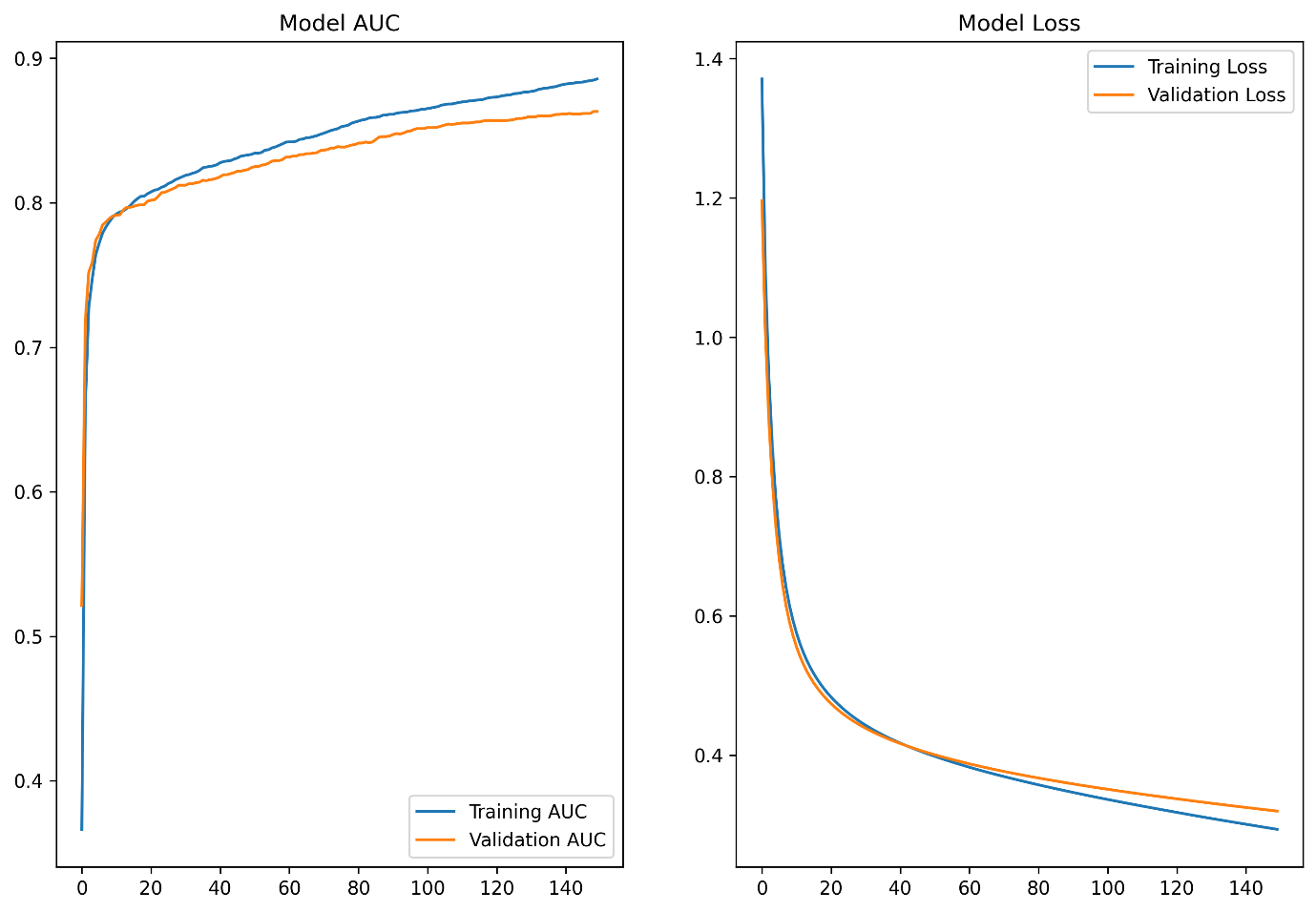
**Results and Analysis**



## ResNet50

## AUC and Loss Curves of ResNet50

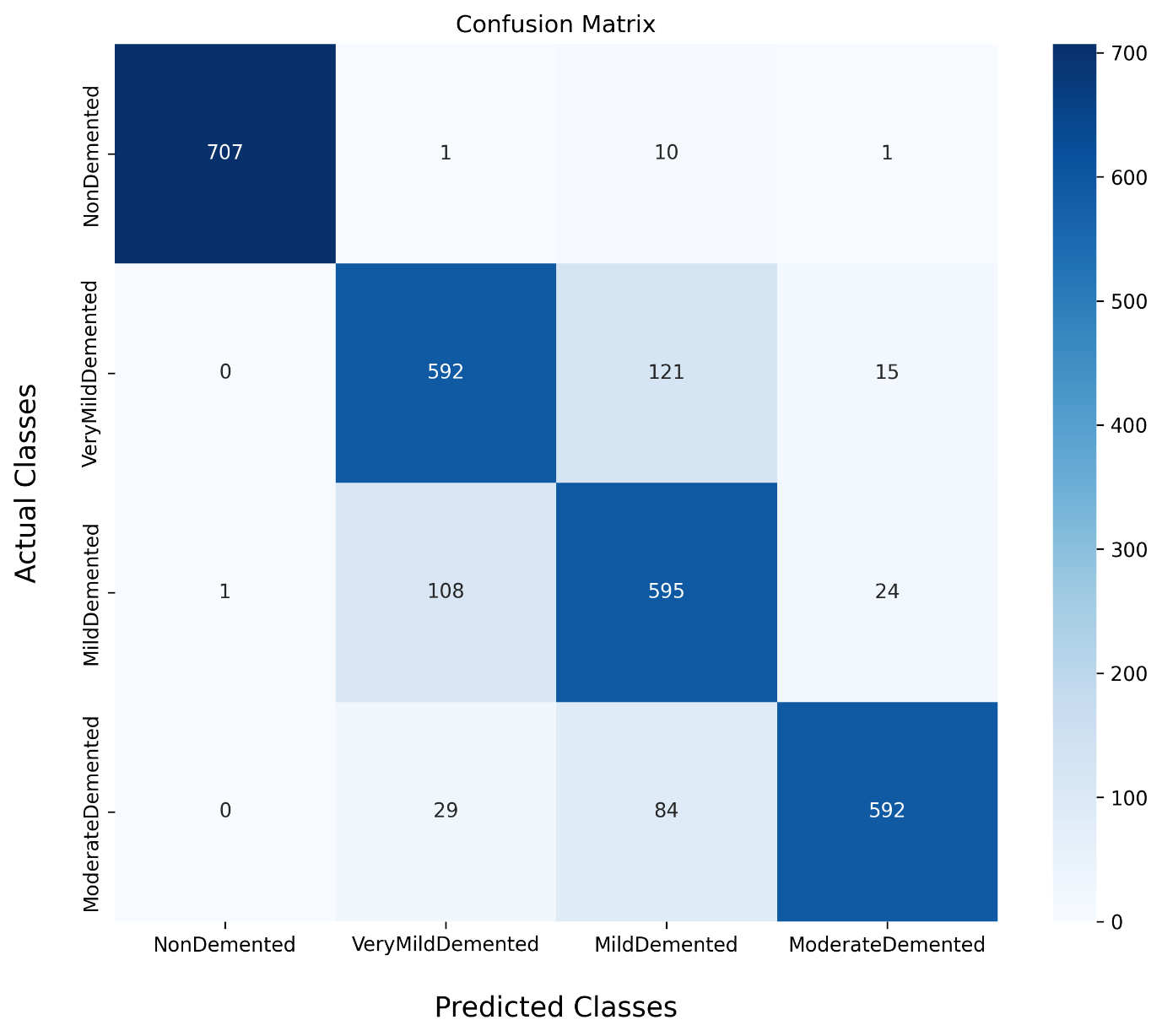
The AUC and loss curves for the ResNet50 model (Figures 7) demonstrate a stable and consistent performance during training. The training AUC gradually increases while the loss decreases, showing good convergence. The validation AUC follows closely, which indicates effective generalization without significant overfitting.



**Figure 9 - ResNet50 Model AUC and Loss Curves**

## Confusion Matrix ResNet50

The confusion matrix for the ResNet50 model (Figure 8) shows high precision and recall for the Non-Demented class, reflecting the model's capability in identifying normal brain conditions. The performance for Very Mild and Mild Demented classes is slightly lower but still consistent, with notable improvements compared to other models. The moderate class shows excellent identification.



**Figure 10 - ResNet50 Confusion Matrix**

## Classification Report of ResNet50

The classification report highlights the model's performance on Alzheimer's disease categories. The **NonDemented** class shows excellent results with a precision, recall, and f1-score of around 99%, indicating very few errors. For the **VeryMildDemented** and **MildDemented** classes, the performance is decent, with f1-scores of 81% and 77%, respectively. The **ModerateDemented** class shows strong precision (94%) but lower recall (84%), leading to an f1-score of 89%. Overall accuracy is **86%**, with balanced performance across the classes but room for improvement in the middle categories.

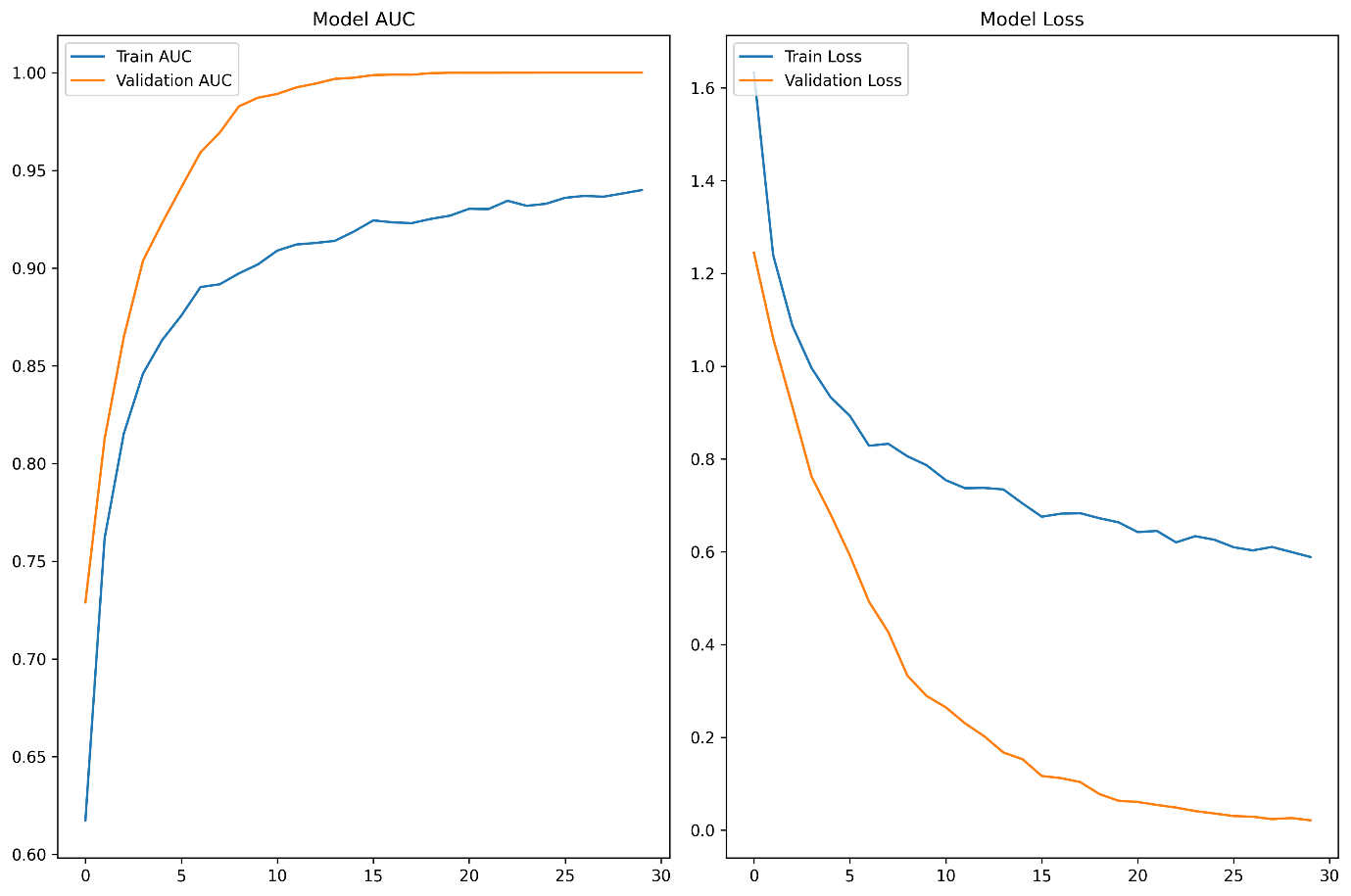
**Table 2 - RESNET50 CLASSIFICATION REPORT**

|  |  |  |  |
| --- | --- | --- | --- |
|  | precision | recall | f1-score |
| NonDemented | 1.0 | 0.98 | 0.99 |
| VeryMildDemented | 0.81 | 0.81 | 0.81 |
| MildDemented | 0.73 | 0.82 | 0.77 |
| ModerateDemented | 0.94 | 0.84 | 0.89 |
| Accuracy (%) | 87 | 86 | 86 |

## DenseNet121

## AUC and Loss Curves of DenseNet121

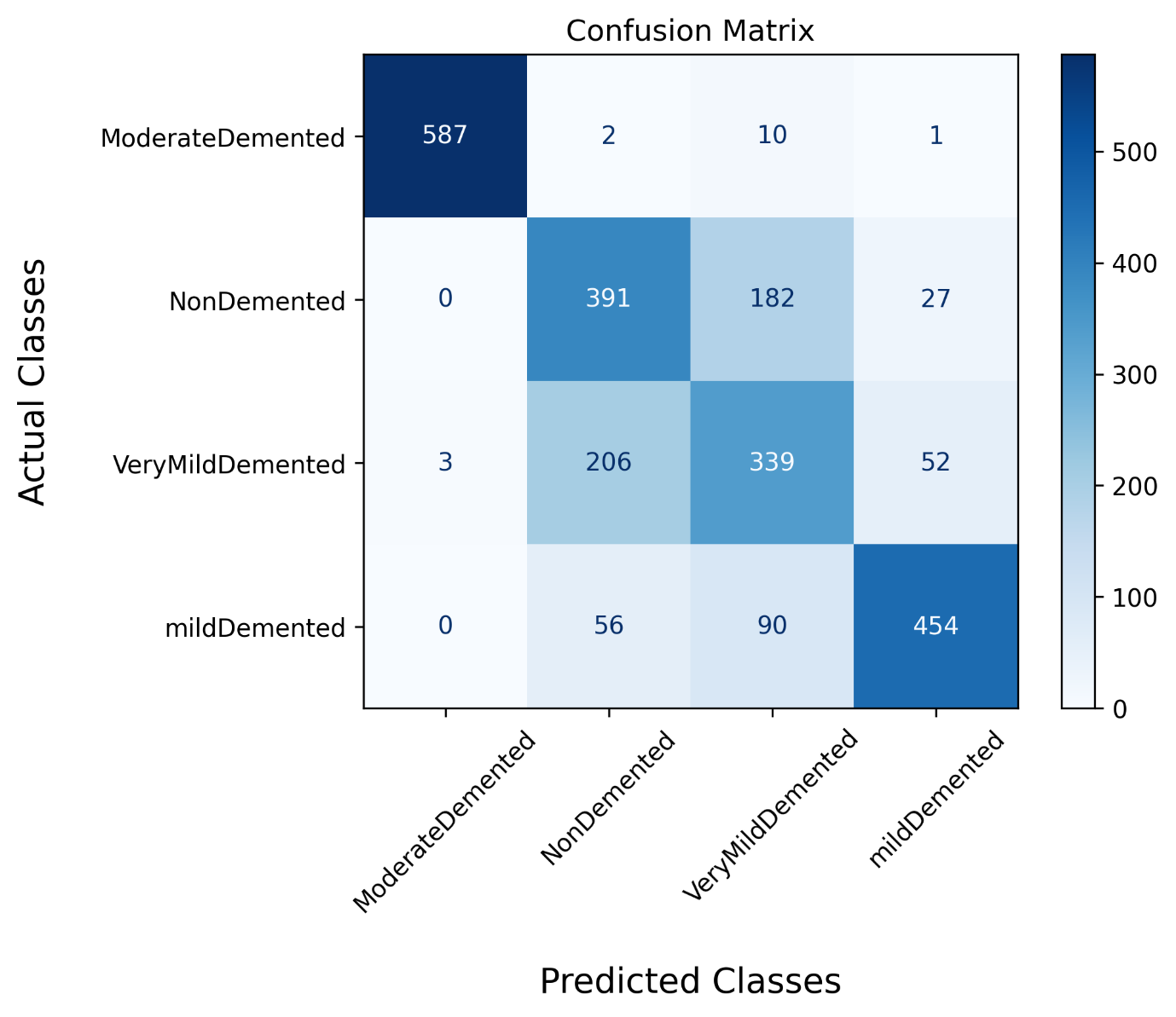
The AUC and loss curves for the DenseNet121 model (Figures 9) show a stable training process, with both training and validation AUC steadily increasing while loss values decrease. This reflects good convergence during training, but the model shows signs of potential overfitting, as the gap between training and validation performance slightly widens.



**Figure 11 - DensNet Model AUC and Loss Curves**

* + 1. **Confusion Matrix DenseNet121**

The DenseNet121 model's confusion matrix (Figure 10) demonstrates excellent classification performance for the ModerateDemented class, achieving nearly perfect results. However, the NonDemented, VeryMildDemented, and MildDemented classes exhibit lower performance. Particularly, the VeryMildDemented class shows some misclassification, as seen from the confusions with other categories.

****

**Figure 12 - DensNet Confusion Matrix**

* + 1. **Classification Report of DenseNet121**

The classification report highlights varied performance across Alzheimer's categories. The **ModerateDemented** class achieved near-perfect precision, recall, and f1-score (~99%). The **NonDemented** class had moderate success with a precision of 65%, recall of 72%, and an f1-score of 69%. Performance for the **VeryMildDemented** class was relatively weaker with an f1-score of 56%. The **MildDemented** class showed stronger performance with a precision of 82% and f1-score of 79%. Overall accuracy was 75%, indicating room for improvement, particularly for the NonDemented and VeryMildDemented classes.

**Table 3 - DENSNET CLASSIFICATION REPORT**

|  |  |  |  |
| --- | --- | --- | --- |
|  | precision | recall | f1-score |
| NonDemented | 0.65 | 0.72 | 0.69 |
| VeryMildDemented | 0.56 | 0.56 | 0.56 |
| MildDemented | 0.82 | 0.76 | 0.79 |
| ModerateDemented | 1.0 | 0.98 | 0.99 |
| Accuracy (%) | 75.75 | 75.50 | 75.75 |

## Comparison of Models (ResNet50 vs. DenseNet121)

ResNet50 achieved an overall test accuracy of 86%, demonstrating robust performance and good generalization ability, as reflected in the smooth curves and minimal gap between training and validation metrics.

In contrast, DenseNet121 exhibited signs of overfitting, as observed from the steep divergence between training and validation loss curves (Figure 8). The model achieved a training accuracy above 90% but struggled during testing with an accuracy of 75%, indicating that the model did not generalize well to unseen data.

## Discussion on Overfitting Issues

The overfitting observed in DenseNet121 can be attributed to excessive complexity and insufficient regularization techniques. Despite high training accuracy, the model underperformed on the test set, highlighting the importance of regularization strategies such as dropout, weight decay, and early stopping.

# Chapter 6

**Conclusion and Future Work**



## Conclusion

This study presented a comparative analysis of two prominent deep learning models, ResNet50 and DenseNet121, for Alzheimer's disease classification based on MRI images. ResNet50 demonstrated strong generalization capabilities, achieving a balanced training accuracy of 88% and a test accuracy of 86%. Its performance across all classes, especially the Non-Demented and Moderate Demented groups, was stable, as confirmed by the confusion matrix. In contrast, DenseNet121, although achieving high training accuracy (above 90%), showed signs of overfitting with a test accuracy of 75%. The model struggled to generalize to unseen data, as evident from the significant divergence between training and validation metrics.

The analysis indicates that while both models hold potential, ResNet50 offers superior performance and is better suited for the Alzheimer’s classification task, based on the current dataset and settings. DenseNet121's overfitting issues suggest that further regularization and optimization strategies are necessary to improve its generalization capability. This study also highlights the importance of balancing model complexity and regularization, especially in medical image classification, where data availability and class imbalances can complicate training.

## Future Work

Although data augmentation helped mitigate the imbalance in the dataset, further improvements are necessary to enhance model accuracy. The current performance suggests that the models still face challenges, which indicates the need for more diverse and original data for better generalization. Future efforts should focus on collecting additional high-quality MRI scans to balance the classes more naturally.

Additionally, exploring advanced models such as transformers and hybrid architectures could provide more robust performance. Techniques like fine-tuning pre-trained models on larger medical datasets or using generative adversarial networks (GANs) to generate synthetic data could also further improve accuracy and reduce overfitting.

# REFERENCES

[1] J. A. S. Lopez, H. M. González, and G. C. Léger, “Alzheimer’s disease,” *Handb. Clin. Neurol.*, vol. 167, pp. 231–255, 2019.

[2] C. Affonso, A. L. D. Rossi, F. H. A. Vieira, and A. C. P. de Leon Ferreira, “Deep learning for biological image classification,” *Expert Syst. Appl.*, vol. 85, pp. 114–122, 2017.

[3] A. Suganya and S. L. Aarthy, “Deep Learning in Medical Image Classification,” in *Integrating Deep Learning Algorithms to Overcome Challenges in Big Data Analytics*, CRC Press, 2021, pp. 139–169. Accessed: Sep. 19, 2024. [Online]. Available: https://api.taylorfrancis.com/content/chapters/edit/download?identifierName=doi&identifierValue=10.1201/9781003038450-9&type=chapterpdf

[4] K. Aziz, “Alzheimer\_CNN.ipynb,” Deep Learning, 2022. [Online]. Available: Alzheimer Multiclass Classification/Alzheimer\_CNN.ipynb

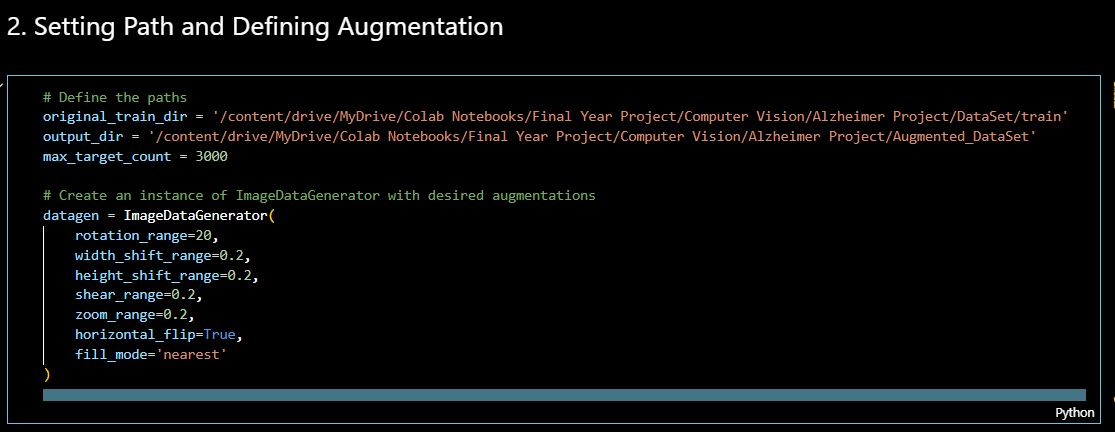
[5] K. Aziz, “Alzheimer\_VGG19.ipynb,” Juypter Notebook, 2022. [Online]. Available: Alzheimer Multiclass Classification/Alzheimer\_VGG19.ipynb

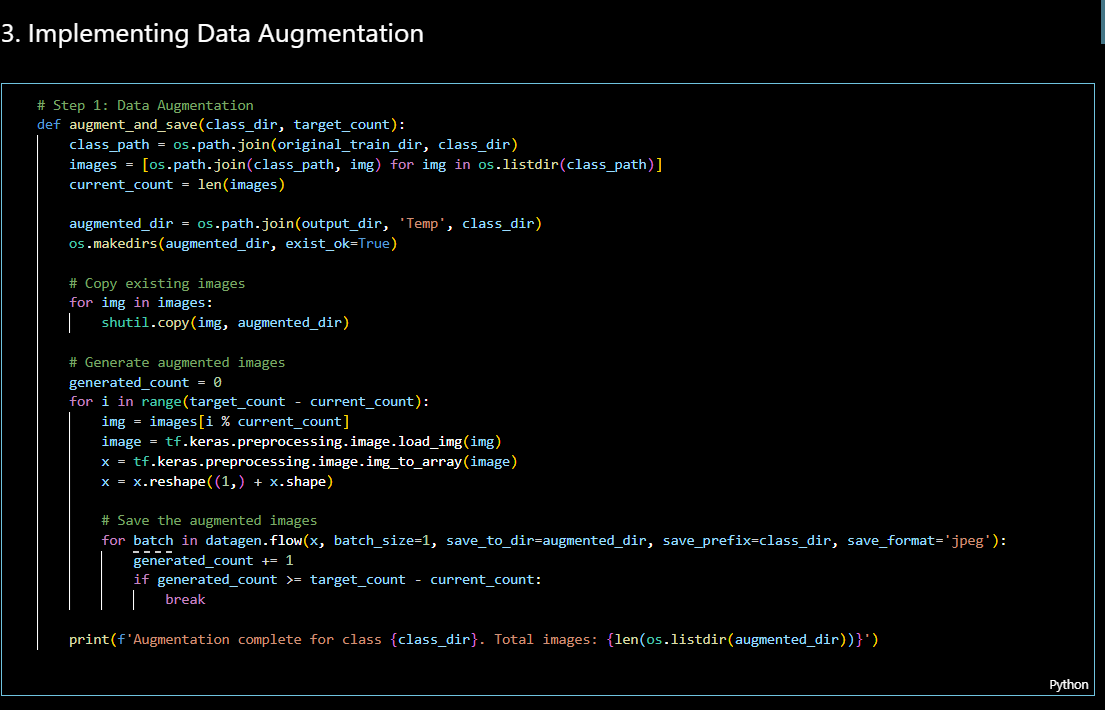
[6] S. Dubey, “Alzheimer’s Dataset ( 4 class of Images).” 2019. [Jpg]. Available: https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images

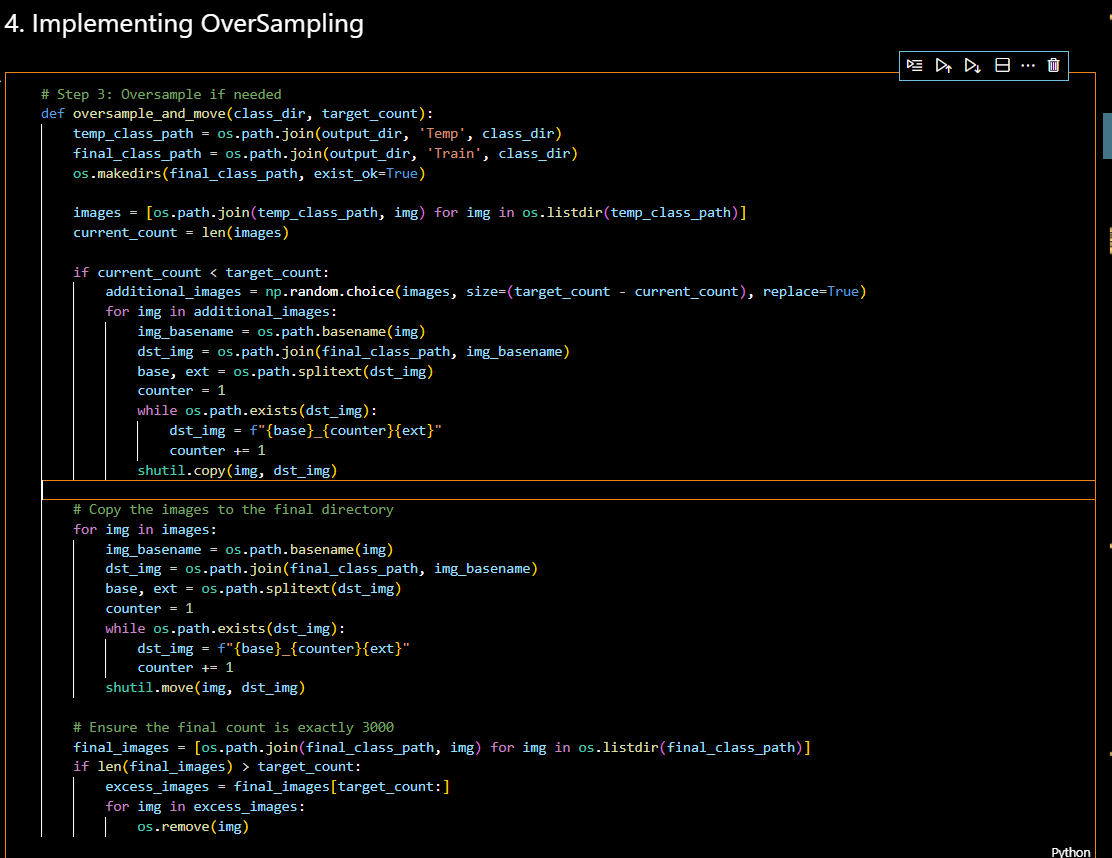
[7] A. Mikołajczyk and M. Grochowski, “Data augmentation for improving deep learning in image classification problem,” in *2018 international interdisciplinary PhD workshop (IIPhDW)*, IEEE, 2018, pp. 117–122. Accessed: Sep. 19, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8388338/

# Appendix A

This section covers the data preprocessing steps performed before training the models. The code handles loading MRI images, splitting the dataset into training and testing sets, applying data augmentation techniques like rotation, zoom, and flipping, and addressing class imbalances. Additionally, it includes steps to resize the images for model compatibility and convert them into arrays for feeding into the DenseNet and ResNet models. The preprocessing ensures that the dataset is optimized for better model performance and effective training across all classes.

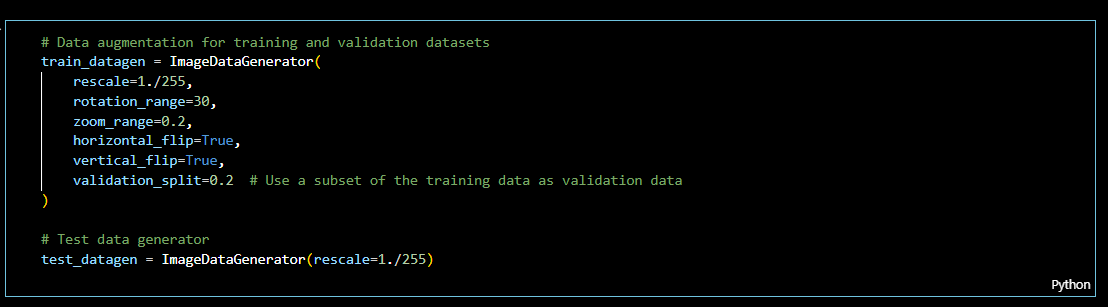


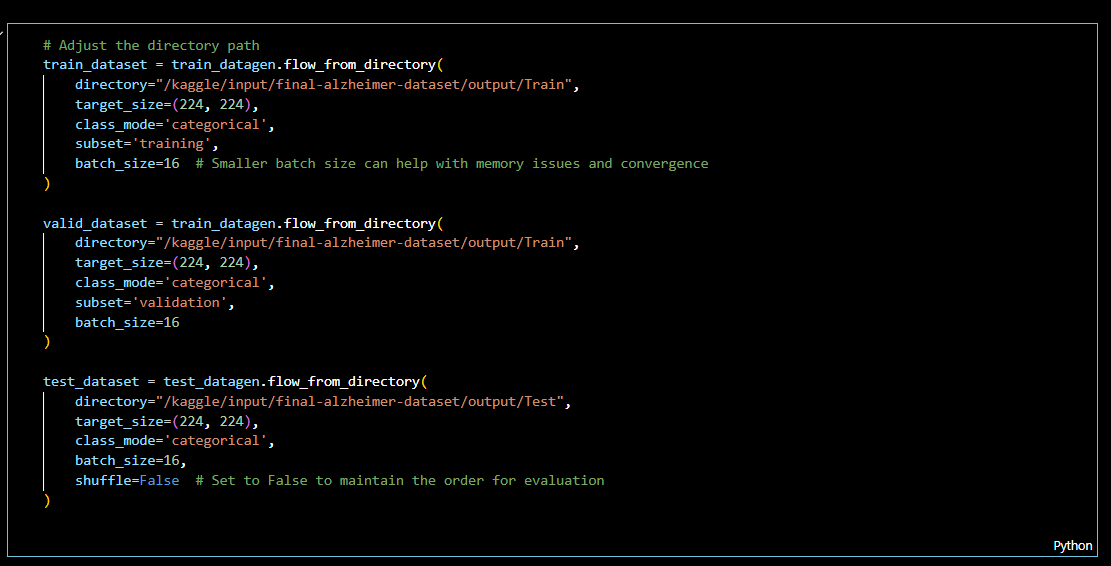




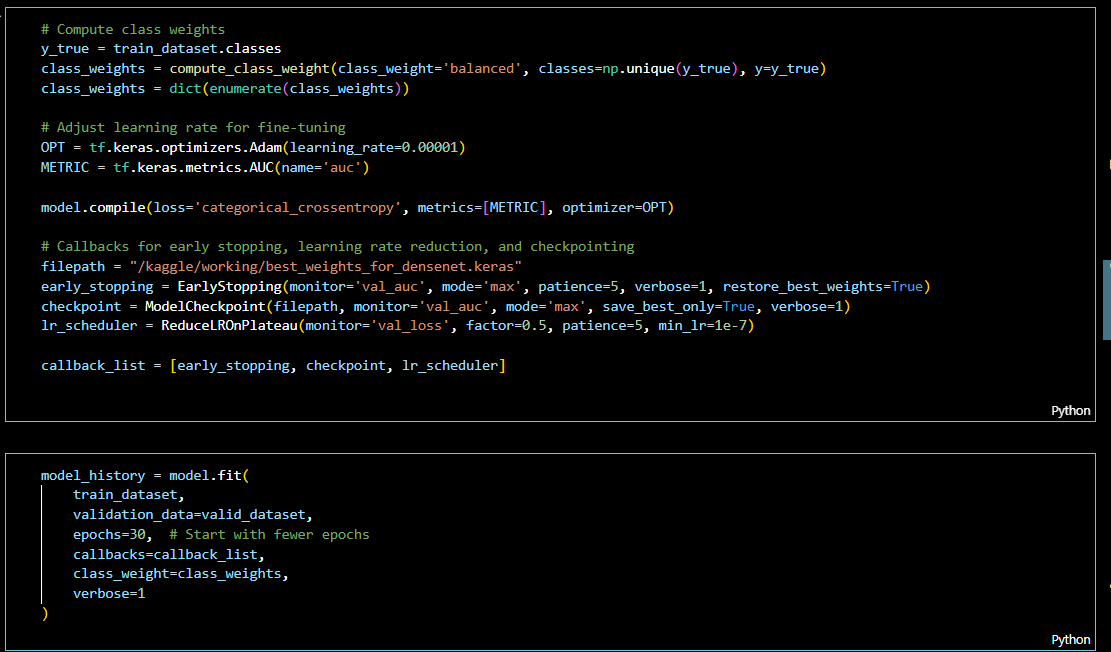
# Appendix B

This section includes the implementation of the DenseNet121 model used for Alzheimer's disease classification. The code begins by loading the pre-trained DenseNet121 model, excluding the top layers to retain the learned features. Custom layers are added on top, including global pooling, dense layers with ReLU activation, and a final softmax output layer for multi-class classification. The model is compiled using the Adam optimizer and categorical cross-entropy loss. The code also includes callbacks for early stopping and checkpoint saving during the training process.









# Appendix C

This section presents the code used for the implementation of the ResNet50 model in the Alzheimer's disease classification task. It utilizes the pre-trained ResNet50 model as a base, with custom top layers for feature extraction. These layers include global average pooling, dense layers with ReLU activation, and a softmax layer for multi-class output. The model is compiled using Adam optimizer with categorical cross-entropy loss. Additionally, the code implements callbacks such as early stopping and learning rate reduction for efficient training.

