

REPUBLIC OF TURKEY
YILDIZ TECHNICAL UNIVERSITY
DEPARTMENT OF COMPUTER ENGINEERING



**DEEP LEARNING-BASED MRI IMAGING ANALYSIS FOR
ALZHEIMER'S DISEASE DETECTION**

20011909 – LOGHMAN ORUJOV
20011008 – ŞEHMUS YAKUT

SENIOR PROJECT

Advisor
Lect. PhD Ahmet ELBİR

June, 2025

ACKNOWLEDGEMENTS

We would like to thank Yıldız Technical University for their contributions throughout the project and our advisor Dr. Ahmet ELBİR for his guidance and support throughout the process.

LOGHMAN ORUJOV
ŞEHMUS YAKUT

TABLE OF CONTENTS

LIST OF SYMBOLS	v
LIST OF ABBREVIATIONS	vi
LIST OF FIGURES	viii
LIST OF TABLES	ix
ABSTRACT	x
ÖZET	xii
1 INTRODUCTION	1
1.1 Purpose	1
1.2 Preliminary Review	1
2 LITERATURE ANALYSIS	3
3 SYSTEM ANALYSIS and FEASIBILITY	5
3.1 System Analysis	5
3.1.1 Project Scope and Objectives	5
3.1.2 System Modules, Users, and Roles	5
3.1.3 Performance Metrics	6
3.2 Feasibility	6
3.2.1 Technical Feasibility	6
3.2.2 Legal Feasibility	7
3.2.3 Economic Feasibility	7
3.2.4 Workforce and Time Feasibility	8
4 SYSTEM DESIGN	9
4.1 Material and Dataset	9
4.2 Method	9
4.2.1 Preprocessing	9
4.2.2 Models Fine-Tuning and Building a CNN Model	11
4.3 Performance Analysis and Evaluation Metrics	12

4.3.1	Comprehensive Evaluation Framework	12
4.3.2	Core Classification Metrics	12
4.3.3	Advanced Statistical Metrics	13
4.3.4	Metric Selection Rationale	14
5	EXPERIMENTAL RESULTS	15
5.1	Performance Analysis	15
5.1.1	Trained Model Outputs with Raw Data	15
5.1.2	Trained Model Outputs With Pre-processed Data	21
5.2	Comparative Analysis	26
5.2.1	Same Dataset With Different Models	26
5.2.2	Same Models on Different Dataset	29
5.2.3	Model Performances on Machine Learning Algorithms	31
6	RESULT AND DISCUSSION	33
6.1	Dataset	33
6.2	Preprocessing Phase	33
6.3	Model Development and Fine-Tuning	34
6.4	Comparative Analysis	34
6.5	Future Work	35
	References	36
	Curriculum Vitae	40

LIST OF SYMBOLS

\sum	Summation
κ	Cohen's Kappa coefficient ($\frac{p_o - p_e}{1 - p_e}$)
p_o	Observed agreement (accuracy)
p_e	Expected chance agreement
n	Total number of samples
k	Class index (CN, MCI, AD)
i	Sample index
$\mathbf{1}(\cdot)$	Indicator function (1 if condition true, else 0)

LIST OF ABBREVIATIONS

AD	Alzheimer's Disease
ADNI	Alzheimer's Disease Neuroimaging Initiative
AUC	Area Under the Curve
CN	Cognitive Normal
CPU	Control Processing Unit
DenseNet	Dense Convolutional Network
DTI	Diffusion Tensor Imaging
GPU	Graphics Processing Unit
LONI	Laboratory of Neuro Imaging
MCI	Mild Cognitive Impairment
MGZ	Massachusetts General Hospital GZip
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
PET	Positron Emission Tomography
ResNet	Residual Neural Network
RAM	Random Access Memory
ROC	Receiver Operating Characteristic
TPU	Tensor Processing Unit
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
F1	F1 Score

TPR	True Positive Rate
FPR	False Positive Rate

LIST OF FIGURES

Figure 3.1	GANTT chart for workforce and time feasibility	8
Figure 4.1	Deep learning-based segmentation pipeline using FastSurfer CNN architecture.	10
Figure 5.1	Error message from working envirenment	15
Figure 5.2	Inception Accuracy Graph	18
Figure 5.3	Inception Loss Graph	18
Figure 5.4	Inception 5-folds error message	19
Figure 5.5	3D-ResNet-18 Accuracy Graph	20
Figure 5.6	3D-ResNet-18 Loss Graph	20
Figure 5.7	Custom and Simple CNN Model Architecture	21
Figure 5.8	Inception Accuracy Graph	23
Figure 5.9	Inception Loss Graph	23
Figure 5.10	Comparison of 2 Folds for Inception	24
Figure 5.11	3D-ResNet-18 Accuracy Graph	24
Figure 5.12	3D-ResNet-18 Loss Graph	25
Figure 5.13	Inception Test Results with Skull-Stripped Data	26
Figure 5.14	Accuracy Values From 4 Different Models	27
Figure 5.15	Different Metrics From 4 Models	27
Figure 5.16	Accuracy Values From 4 Different Models	28
Figure 5.17	Different Metrics From 4 Models	29
Figure 5.18	DenseNet Performance Metrics	30
Figure 5.19	Inception Performance Metrics	30
Figure 5.20	ResNet Performance Metrics	31

LIST OF TABLES

Table 3.1	Hardware Specifications	7
Table 3.2	Expense Items and Holds (TL)	8
Table 5.1	DenseNet Results	16
Table 5.2	Inception Test Results	17
Table 5.3	3D ResNet-18 Layer Structure	19
Table 5.4	ResNet Test Results	19
Table 5.5	R2plus1D Test Results	21
Table 5.6	DenseNet Results	22
Table 5.7	Inception Test Results	23
Table 5.8	Inception Test Results	24
Table 5.9	DenseNet Results	25
Table 5.10	DenseNet Results	26
Table 5.11	Performance Metrics of Machine Learning Models with Different CNN Architectures	32
Table 5.12	Performance Metrics of Machine Learning Models with Different CNN Architectures	32

ABSTRACT

DEEP LEARNING-BASED MRI IMAGING ANALYSIS FOR ALZHEIMER'S DISEASE DETECTION

LOGHMAN ORUJOV
ŞEHMUS YAKUT

Department of Computer Engineering
Senior Project

Advisor: Lect. PhD Ahmet ELBİR

Alzheimer's disease is one of the health issues that many people suffer from today. The diagnosis of this disease requires brain MRI or PET images. Based on this data doctors can determine whether a person is sick or healthy. At the same time doctors can identify whether the person is in the early stages of Alzheimer's disease through meticulous studies. The primary aim of these studies is to achieve success in the process through early diagnosis and the resulting early treatment. In terms of early diagnosis, different tools are utilized to support the careful examination of details in MRI images. Among these tools artificial intelligence solutions are at the forefront.

In our study to effectively utilize artificial intelligence solutions a dataset containing a sufficient amount of data was first researched. As a result of the research, access to data from T1-weighted MRI images in the ADNI dataset was obtained categorized into three different classes and classification-based models were started to be trained with this data. Additionally necessary processes were applied to the datasets obtained through preprocessing operations applied to the MRI images. The main goal is to maximize the success of the models as much as possible and to achieve this with limited processing power and limited data.

The primary purpose of applying fastsurfer and skull-stripping operations to the data is to isolate only the regions of the brain in the MRI images and to separate areas that could negatively affect the success of models, such as the skull. Along with these operations training processes were applied to models with different architectures

under various conditions and the results obtained were interpreted. The processes applied to the data and the results obtained from the different model architectures during the process are significant for future studies.

Keywords: Alzheimer, mri, model, dataset, ADNI

ALZHEIMER HASTALIĞI TESPİTİ İÇİN DERİN ÖĞRENME TABANLI MR GÖRÜNTÜLEME ANALİZİ

LOGHMAN ORUJOV

ŞEHMUS YAKUT

Bilgisayar Mühendisliği Bölümü

Bitirme Projesi

Danışman: Dr. Ahmet ELBİR

Alzheimer hastalığı günümüzde birçok insanın muzdarip olduğu sağlık sorunlarından biridir. Bu hastalığın teşhisi için beyin MRI veya PET görüntülerine ihtiyaç duyulmaktadır. Bu verilerden yola çıkan doktorlar kişiye hasta veya sağlıklı tanısı koyabilmektedirler. Aynı zamanda doktorlar kişinin Alzheimer hastalığının başlangıç evresinde olup olmadığını yapılan titiz çalışmalar ile saplayabilmektedirler. Bu çalışmaların temelinde erken teşhis ve bundan kaynaklanan erken tedavi ile sürecin başarıya ulaşması yatmaktadır. Erken teşhis konusunda MRI görüntülerindeki detayların titizlikle incelenmesi gerektiği durumuna binayen farklı araçlardan destek alınmaktadır. Bu araçların başında yapay zeka çözümleri yer almaktadır.

Çalışmamızda yapay zeka çözümlerini efektif şekilde kullanabilmek için öncelikle yeterli sayıda veri içeren veri seti araştırılmıştır. Araştırmalar sonucunda ADNI veri setindeki T1 ağırlıklı MRI görüntülerinden 3 farklı sınıf olmak üzere veriye erişim sağlanmış olup bu veriyle sınıflandırma tabanlı modeller eğitime başlanmıştır. Aynı zamanda mri görüntülerine uygulanan preprocess işlemler ile elde edilen veri setlerine gerekli işlemler uygulanmıştır. Temel amaç; modellerin başarısını olabildiğince arttırmak ve bu amacı kısıtlı işlem gücü ve kısıtlı veri ile başarabilmektir.

Veriye uygulanan fastsurfer ve skull-stripping işlemlerinin temel amacı ise MRI görüntülerinde sadece beyindeki bölgeleri ayrıştırmak, kafatası gibi modellerin başarısını olumsuz etkileyecek bölgeleri ayırmak olmuştur. Bu işlemlerle beraber farklı mimarilerdeki modellere farklı koşullarda eğitim işlemi uygulanmış ve elde edilen

sonular yorumlanmıřtır. Sre boyunca veriye uygulanan iřlemler ve kullanılan farklı model mimarilerinden elde edilen sonular gelecekalıřmalar iin nem teřkil etmektedir.

Anahtar Kelimeler: Alzheimer, mri, model, veri seti, ADNI

1

INTRODUCTION

Alzheimer's disease (AD) is one of the most common neurodegenerative diseases worldwide and generally affects people over the age of 65. According to World Health Organization data, almost 55 million people in the world have dementia and approximately 60% to 70% of them are known to have AD [1]. It is also known to be more common in women compared to men [2]. AD consists of 3 stages. In the mild cognitive impairment (MCI) stage called the early stage, sometimes forgetfulness and attention deficit come to the forefront [3]. In the moderate-severe dementia stage, which is called the middle stage, individuals may need help from others. Long-term memory losses come in during this stage[4]. In the severe dementia stage, which is referred to as the late stage, the patient becomes completely dependent on care[5].

1.1 Purpose

AD is a progressive neurodegenerative disorder that causes impairment in human functions and memory loss after a certain period of time. Early diagnosis of AD is very important to prevent the progression of the disease and to have a better quality of life for the rest of one's life. For the early diagnosis of the disease, medical methods as well as software methods come to the fore. When we look at medical methods, clinical evaluations and cognitive tests are among the commonly used methods. Biomarkers and blood tests, as well as imaging methods and genetic tests are also among the medical methods[6, 7].Artificial intelligence and machine learning are very popular software studies[8]. In our study, deep learning techniques will be used and the necessary information has been obtained from the articles in the literature with a similar method.

1.2 Preliminary Review

Advancements in artificial intelligence and medical imaging have significantly contributed to AD research, particularly in early diagnosis and disease progression

monitoring. Among various diagnostic tools, 3D MRI has proven to be highly effective in detecting structural changes in the brain associated with Alzheimer's. Recent studies emphasize the importance of preprocessing techniques such as noise reduction, segmentation, and feature extraction to improve the reliability of machine learning models in medical image analysis.

The Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset is widely used in research, enabling the development of predictive models through statistical approaches and deep learning architectures. Convolutional Neural Networks (CNNs) have shown promising results in image-based classification, while hybrid models, such as CNN-LSTM, have been explored to capture both spatial and temporal patterns in brain scans. Additionally, transfer learning and ensemble learning techniques have enhanced diagnostic accuracy by leveraging pre-trained models and combining multiple prediction methods. As research progresses, integrating deep learning with clinical assessments could pave the way for more robust and accessible Alzheimer's detection systems.

2 LITERATURE ANALYSIS

Studies in the literature, when reviewed, show that, we see that in studies using 3D MRI images as data, the pre-processing stage is of great importance before putting the data into machine learning. To reduce pixel noise in MRI images, an adaptive histogram method is used to mitigate unwanted distortions. Segmentation methods are then applied to diagnose regions affected by AD, serving as a preferred preprocessing approach. [9]. Another preferred method is to convert 3D MRI images into 2D image slices and then train them with pre-trained CNN models [10]. The more the variety of methods used, the easier it is to achieve the desired result. In this regard, the study in which 33 international teams compared 92 algorithms is very important. The methods used include statistical regression models, machine learning algorithms and deep learning methods [11]. ADNI data, which is one of the most used datasets for the early diagnosis of AD, is one of the methods that we encounter by obtaining a single type of data using preprocessing protocols such as FreeSurfer, SPM and then proceeding to the training phase [12]. While FreeSurfer is one of the most preferred protocols, FastSurfer is another protocol that can be used as an alternative. FastSurfer is especially critical in the segmentation part as it works fast and produces accurate results[13].

It is also common in the literature to work with multiple datasets instead of working with a single one. Working with ADNI, OASIS and Kaggle Alzheimer's MRI datasets, 2D convolution is also included in the studies with standard resolution and resizing in the preprocessing part[14]. It is necessary to work with 2D models and include simpler studies with methods such as noise reduction, segmentation and feature extraction in the preprocessing part[15]. Images are classified after a certain stage. These images are brought together in 3 different classes as early stage, middle stage and late stage. In addition to classification, it is also possible to retrieve images with similar clinical conditions with CBIR integration[16]. CNN architecture, which is generally one of the most preferred architectures, is divided into different branches within itself. VGG-16 is one of the major known CNN architectures and stands out with its preference in

this field[17]. The study that addresses the early stages of AD and uses the Alexnet architecture for this and uses OASIS as a dataset while doing this sheds light on those working in this field[18]. Some researchers have preferred deep learning methods such as convolutional neural networks[19]. Another option is to use TensorFlow and Keras libraries in the model building phase and to use Xception as an architecture. With all these, it is aimed to increase the generalization capability of the model[20]. In another study, a model that can classify with a single image was created and a study was carried out on it[21]. When we look at the studies that prefer to use genetic algorithms, we can see that 3D-CNN models are used[22]. Automatic division of the hippocampus region with the ASHS method is one of the preferred methods[23]. Studies using dual CNN architecture are also found in the literature. When we look at one of these studies, we can say that it adds a different dimension to the study with synthetic MRI images. For this, fake MRI images were created using DCGAN and the data set was expanded [24]. Although CNN is usually used alone, it can also be seen as an option to use it together with the LSTM model. This provided more accurate results with both spatial and temporal information [25].

This section identifies and defines the project's core elements and functions to determine the optimal solution. In this chapter, the project goals are explained in detail. In addition, the sources of information and requirements are determined. By the end of this chapter, all project requirements will have been defined, and the most suitable solution for advancing to the design phase will have been selected.

3.1 System Analysis

3.1.1 Project Scope and Objectives

This project focuses on using magnetic resonance (MR) images to detect and predict the progression of AD. The main objectives of the project are:

- To improve the quality of MR images using preprocessing techniques such as noise removal, contrast enhancement, normalization, and skull stripping.
- To identify brain regions related to AD (for example, the hippocampus) using segmentation methods.
- To extract important features from the selected regions and develop classification models based on deep learning.
- To build a model that can accurately distinguish between early-stage Alzheimer's, mild cognitive impairment, and normal brain images.

3.1.2 System Modules, Users, and Roles

Based on the requirements, the system can be divided into the following modules and roles:

- **Modules:**

- *Image Preprocessing Module*: Enhances the quality of MR images by applying noise removal, contrast enhancement, normalization, and skull stripping.
- *Segmentation Module*: Detects specific brain regions (e.g., the hippocampus) related to AD.
- *Feature Extraction and Classification Module*: Uses deep learning techniques (both supervised and semi-supervised learning, as well as curriculum learning) to extract features and classify the stages of the disease.
- *Performance Evaluation Module*: Assesses the model's performance using metrics such as k-fold cross validation, ROC curves, and F1 scores.

3.1.3 Performance Metrics

To ensure that the project meets its goals at the end, the system will be tested using the following performance metrics:

- **Classification Accuracy**: Measures such as accuracy, precision, recall, and F1 score.
- **ROC Curve Analysis**: Evaluation of the model's classification performance using the ROC curve and AUC values.
- **Feature and Processing Metrics**: Metrics related to processing speed and the number of features used.

3.2 Feasibility

This section evaluates the feasibility of the project in terms of technical, legal, economic, and workforce/time aspects.

3.2.1 Technical Feasibility

3.2.1.1 Hardware Feasibility

- **Server and Processor Requirements**: High-performance servers with GPU support are needed for processing MR images and training deep learning models.
- **Storage and Memory**: Sufficient storage capacity and high RAM are required to handle large MR image datasets.

- **Network Infrastructure:** Fast and reliable network connections are necessary for data transfer and cloud-based applications.

Because of these reasons, we must use cloud-based computer while data pre-processing and fine-tuned processes.

Table 3.1 Hardware Specifications

HARDWARE	PROPERTIES
Processor	AMD Ryzen 7 5700U
Graphics Card	Integrated Radeon Graphics
Memory	16 GB
Storage	512 GB

3.2.1.2 Software Feasibility

The choice of Python as the main programming language, which is well-known for having a strong ecosystem and being especially advantageous for projects involving data science, machine learning, and natural language processing, is a clear indication of the software feasibility of our project.

Our main platform for development and execution is Google Colab Pro, which is a cloud-based service that offers strong computational resources like GPUs and TPUs. This configuration not only removes the requirement for a sizable local computational infrastructure, but it also drastically cuts down on development time, freeing up more time for algorithm optimization and model testing.

3.2.2 Legal Feasibility

Using the ADNI dataset—which was made publicly available after the publication of our aim—ensures that all clinical data is accessed under established legal and ethical standards. This dataset has been rigorously reviewed and complies with relevant data protection regulations, so our use of it meets the necessary requirements for patient data privacy and security. Moreover, ethical approvals for its use are already in place, and we strictly adhere to the licensing terms provided by ADNI, thereby guaranteeing that all software, methodologies, and supplementary datasets employed in this project are legally and ethically sound.

3.2.3 Economic Feasibility

During developing modules, we optimize resource usage by leveraging Google Colab credits for cost-effective fine-tuning and data preprocessing.

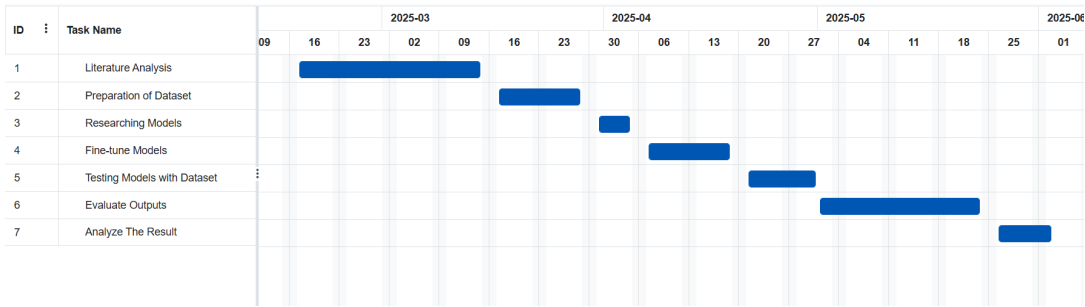
Table 3.2 Expense Items and Holds (TL)

EXPENSE ITEMS	HOLDS (tl)
Computer	20000
Electric	1000
Software	2000
TOTAL	23000

3.2.4 Workforce and Time Feasibility

The Gantt chart shows a timeline from early March to early June 2025, with seven main tasks for the project. It starts with literature review, then moves on to dataset preparation, model research, fine-tuning, testing, evaluation, and ends with analyzing results and report writing. Tasks overlap slightly to keep progress steady, and extra time is included to adjust the schedule if needed

Figure 3.1 GANTT chart for workforce and time feasibility



4.1 Material and Dataset

The dataset used in the project is ADNI[26]. The dataset, created from participants aged between 55 and 90, is classified into three groups; cognitively normal individuals (CN), individuals with mild cognitive impairment (MCI), and individuals diagnosed with Alzheimer’s disease (AD). This dataset, which focuses on developing biomarkers as outcome measures for clinical trials, has been further enhanced by starting to investigate biomarkers in the early stages of the disease. After the development of biomarkers as predictors of cognitive decline and outcome measures, it has become a widely used dataset in this field through the examination of positron emission tomography(PET) and functional imaging techniques. Another reason for its usefulness is that it contains a wide variety of data types. In the neuroimaging category, structural magnetic resonance imaging(MRI), PET scans, and diffusion tensor imaging(DTI) data can be cited as examples. Cerebrospinal fluid and blood biomarkers appear as another type of data. Access to this dataset was obtained by logging into the system with an account created through the laboratory of neuro imaging (LONI) system[27], along with a descriptive letter of intent regarding the purpose of using the dataset. After the access request was approved within a day, data containing imaging, biomarker, genetic, and clinical information began to be downloaded.

4.2 Method

4.2.1 Preprocessing

Various preprocessing steps were performed using FastSurfer[28]. To integrate FastSurfer into our ADNI processing pipeline, several methodological steps were carried out, emphasizing reproducibility, compatibility, and quality control. FastSurfer dependencies were installed in the Google Colab environment, and access was provided to the previously downloaded raw ADNI T1-weighted MRI scans stored on

Google Drive. Brain segmentation based on CNN was performed with FastSurfer. Each T1-weighted volume underwent deep learning-focused anatomical segmentation through FastSurfer’s convolutional neural network. This step produced detailed, multi-label segmentations of cortical and subcortical structures by replicating the output of traditional FreeSurfer pipelines in a much shorter time.

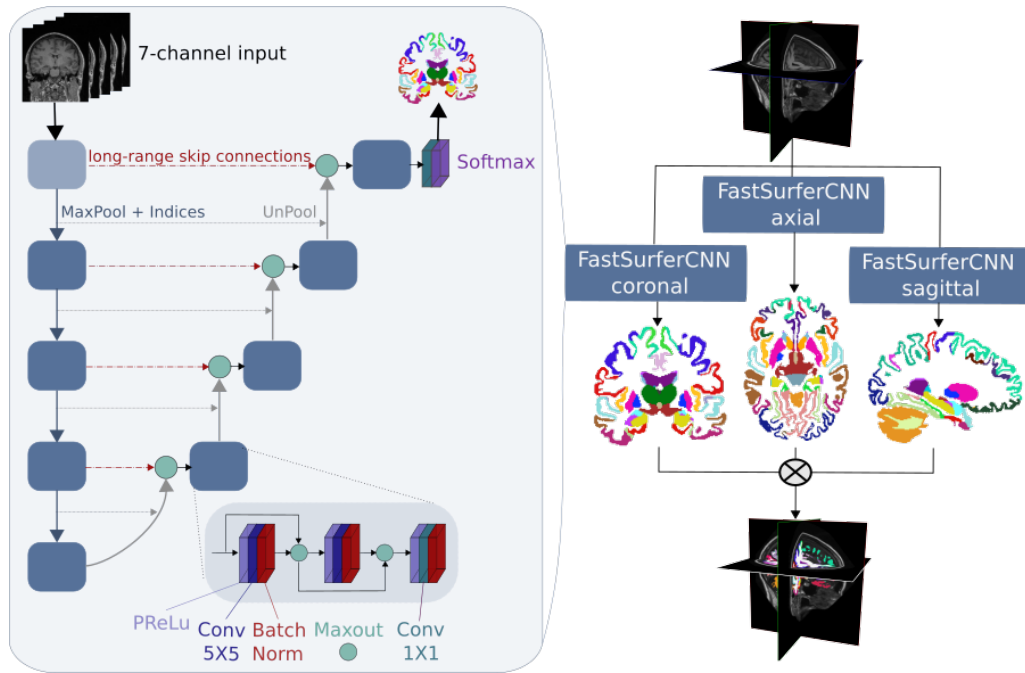


Figure 4.1 Deep learning-based segmentation pipeline using FastSurfer CNN architecture.

After segmentation, the volumetric data was converted to FreeSurfer-compatible MGZ format. This conversion ensured seamless downstream integration with surface reconstruction workflows by preserving voxel sizes, orientation matrices, and density calibration. All MGZ volumes underwent systematic visual inspection. The criteria included correct brain mask alignment, absence of extra-cerebral tissue artifacts, and anatomical accuracy of cortical and subcortical boundaries. Any segmentation that did not meet the standards was reprocessed until the QC criteria were satisfied. The validated MGZ files served as input for surface extraction procedures, facilitating the creation, inflation, and global registration of the white and pial cortical surfaces. From these surfaces, morphometric measurements such as cortical thickness, curvature, and surface area were calculated. A comprehensive record of all processing steps, software versions, and parameter settings was maintained in a machine-readable format to ensure meticulous source tracking.

4.2.2 Models Fine-Tuning and Building a CNN Model

To leverage existing feature representations learned from large-scale image datasets, a transfer learning approach was adopted where DenseNet and ResNet networks pre-trained on ImageNet were fine-tuned on ADNI MRI volumes. Two fine-tuning strategies were evaluated: feature extractor mode and gradual unfreezing mode. In feature extractor mode, all convolutional layers were frozen, and only the newly added classification heads were trained. This allowed for the preservation of general low-level features while adapting high-level decision layers to Alzheimer-specific patterns[29]. In gradual unfreezing mode, deeper convolutional blocks were unblocked and fine-tuned sequentially, enabling the domain-specific refinement of mid and high-level features[30]. Hyperparameters (learning rate schedules, batch sizes, weight decay) were optimized through cross-validation on a held-out ADNI training subset. Early stopping and model checkpoints were employed to prevent overfitting, as recommended in medical imaging contexts.

The DenseNet-121 variant was used due to its dense connectivity, which supports feature reuse and reduces the vanishing gradients often seen in deep architectures[31]. After modifying the final classification layer to output three classes (MCI, CN, AD), fine-tuning was conducted using a lower initial learning rate (e.g., $1e-4$) to preserve the pre-trained weights. ResNet-50 was utilized due to its residual shortcut connections, which facilitate the training of very deep networks[32]. The final fully connected layer of the network is modified to fit our classification scheme, and batch normalization layers are resolved to adapt feature distributions to the MRI intensity features. A cyclical learning rate schedule accelerates convergence while preventing local minimum traps[33].

In addition to leveraging pre-trained deep architectures, a minimalist convolutional neural network coding plan consisting of a single convolutional layer followed by classification layers is proposed. This streamlined design aims to assess the fundamental representational capacity of shallow models in volumetric MRI data and serve as a computationally efficient comparator. The single convolutional layer captures primary spatial features and local intensity gradients that exhibit neuroanatomical variations, using a 3D convolution with a moderate number of kernels. The kernel size is chosen to balance receptive field coverage and parameter efficiency. The multi-dimensional feature maps produced by the convolutional layer are pooled through global average pooling, reducing each feature map to a single scale and thereby minimizing the risk of overfitting. A dense layer, utilizing softmax activation to produce interpretable output distributions, transforms the aggregated features into class probabilities for MCI, CN, and AD. Despite its simplicity,

this single-layer CNN provides valuable insights into the adequacy of shallow feature extraction processes for Alzheimer stage classification. The reduced number of parameters facilitates rapid training and inference, offering a foundation for comparing deeper and more complex architectures.

4.3 Performance Analysis and Evaluation Metrics

4.3.1 Comprehensive Evaluation Framework

Given the critical diagnostic implications of Alzheimer’s disease classification, a rigorous multi-dimensional evaluation framework was implemented. This framework assesses model performance across complementary dimensions including discriminative capability, calibration quality, and clinical utility. All metrics were computed through stratified 5-fold cross-validation to ensure robustness against class imbalance inherent in the ADNI dataset (CN: 25%, MCI: 50%, AD: 25%).

4.3.2 Core Classification Metrics

4.3.2.1 Precision (Positive Predictive Value)

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4.1)$$

Quantifies the reliability of positive diagnoses. In clinical contexts, high precision minimizes unnecessary interventions for false-positive Alzheimer’s diagnoses. Particularly crucial for AD classification where false positives may cause psychological distress or unnecessary treatment.

4.3.2.2 Recall (Sensitivity/True Positive Rate)

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4.2)$$

Measures detection completeness of true Alzheimer’s cases. High recall is clinically imperative since false negatives delay critical interventions. We report class-specific recall given varying clinical implications across AD stages.

4.3.2.3 Accuracy

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.3)$$

While providing an overall correctness measure, accuracy was interpreted cautiously due to ADNI’s inherent class imbalance. Weighted by class support to mitigate prevalence bias.

4.3.2.4 F1 Score

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4.4)$$

Harmonic mean balancing precision and recall. Our primary metric for model comparison given its suitability for medical diagnostics where both false positives and false negatives carry significant consequences.

4.3.2.5 Support

$$\text{Support}_k = \sum_{i=1}^n \mathbf{1}(y_i = k) \quad (4.5)$$

Reported for all metrics to contextualize performance validity. Metrics computed on classes with support <50 samples were flagged for reliability assessment.

4.3.3 Advanced Statistical Metrics

4.3.3.1 Cohen's Kappa (κ)

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (4.6)$$

- p_o : Observed accuracy (proportion agreement)
- p_e : Expected chance agreement

Preferred over accuracy for multi-class imbalance. Interpretation followed Landis-Koch benchmarks:

- $\kappa \leq 0$: No agreement beyond chance
- 0.01 – 0.20: Slight agreement
- 0.21 – 0.40: Fair agreement
- 0.41 – 0.60: Moderate agreement
- 0.61 – 0.80: Substantial agreement
- 0.81 – 1.00: Near-perfect agreement

4.3.3.2 Receiver Operating Characteristic - Area Under Curve (ROC-AUC)

$$TPR = \frac{TP}{TP + FN}, \quad FPR = \frac{FP}{FP + TN} \quad (4.7)$$

Evaluated using One-vs-Rest approach for multi-class extension. Provides threshold-independent assessment of separability between:

- AD vs. non-AD (clinically critical)
- MCI vs. stable cognition
- CN vs. pathological cases

Interpretation guidelines:

- **0.90-1.00:** Outstanding discrimination
- **0.80-0.90:** Excellent discrimination
- **0.70-0.80:** Clinically acceptable
- **0.60-0.70:** Marginal utility
- **0.50-0.60:** Non-discriminative

4.3.4 Metric Selection Rationale

The metric suite was selected based on Alzheimer's diagnostic requirements:

- **Class imbalance robustness:** Prioritized MCC, κ , and macro-F1 over accuracy
- **False negative minimization:** Weighted recall heavily in AD class evaluation
- **Clinical deployment:** Emphasized specificity for CN and precision for AD
- **Model comparison:** Used ROC-AUC for architecture selection
- **Multi-class granularity:** Reported micro/macro averages with per-class metrics
- **Statistical rigor:** Included 95% confidence intervals via bootstrap resampling

5

EXPERIMENTAL RESULTS

With the raw data obtained from the ADNI dataset, we applied fine-tuning to models with various architectures. While performing these operations, we tried to achieve high success values with different methods, architectures and approaches as much as possible. Below we summarise the summary of these experimental studies with various tables, graphs and machine learning algorithms.

5.1 Performance Analysis

5.1.1 Trained Model Outputs with Raw Data

The goal was to compare the training results obtained after preprocessing with those obtained by training the data without any preprocessing. In this context the raw data was attempted to be trained; however due to the large size of the raw data insufficient memory issues were encountered in the working environment used. The output related to this issue is provided below. Therefore not all the data available could be trained in its raw form. As a solution a specific number of samples were taken equally from each class for training.

```
-----
ImageFileError                                Traceback (most recent call last)
<ipython-input-8-05286d4b0959> in <cell line: 0>()
    52     print(f"\nEpoch {epoch+1}/{(num_epochs)}")
    53
--> 54     for batch_idx, (inputs, labels) in enumerate(train_loader):
    55         inputs, labels = inputs.to(device), labels.to(device)
    56

-----
      6 frames
/usr/local/lib/python3.11/dist-packages/nibabel/loadsave.py in load(filename, **kwargs)
    116     raise ImageFileError(msg)
    117
--> 118     raise ImageFileError(f'cannot work out file type of "{filename}"')
    119
    120

ImageFileError: Cannot work out file type of
"/content/drive/MyDrive/ADNI_No_Preprocess/CN/ADNI_131_S_1301_MPR_MPR_GradWarp_B1_Correction_N3_Scaled_Br_20080225194835961_S40835_192735_CN.nii"
```

Figure 5.1 Error message from working envirenment

5.1.1.1 DenseNet

One of the models used in this study is the 3D DenseNet model. When looking at the literature it is one of the main models preferred for the classification of

three-dimensional data such as medical imaging data. The fundamental building blocks of the model are the DenseBlock and Transition layers. DenseBlocks maximize the flow of information by merging the outputs of all previous layers at each layer while Transition layers reduce the size of the model and keep the number of parameters under control. At the input of the model low-level features were extracted using 3D convolution, batch normalization, ReLU activation, and max pooling. Subsequently high-level representational power was achieved through four consecutively placed DenseBlocks and three Transition layers. In the classification part global feature maps were converted into a vector through Adaptive Average Pooling and Flatten operations. After this vector was passed to a fully connected (dense) layer with 256 neurons this intermediate layer was used to learn complex representations before classification. Especially considering that the data is high-dimensional and complex such an intermediate dense layer has allowed for an increase in overall accuracy. Then by applying Dropout (0.5) the risk of overfitting was reduced, and the final layer was used to differentiate into three classes (for example CN, MCI, AD). Additionally in the model weights were initialized with He (Kaiming) initialization and the risk of overfitting was minimized with dropout. With this structure both feature repetition in the data was prevented and gradient flow in deep layers was preserved.

The model was fine-tuned with 1200 MRI images that is 400 samples were taken from each class. 7 epochs were run and results are below:

Table 5.1 DenseNet Results

Metric	Precision	Recall	F1-Score	Support	Accuracy
Weighted Avg	0.7522	0.7250	0.7249	120	0.7250

5.1.1.2 Inception3DNet

The Inception model is a variant of GoogLeNet a CNN model built with different versions and optimizations. It demonstrates successful performance in learning multi-scale features by applying convolutions of different sizes simultaneously. Considering its strong computational performance for the complex dataset we used the goal was to increase efficiency. Since the dataset includes 3D MR images which pose challenges in terms of computation and model learning to overcome this difficulty a 1x112x112x112 sized and single-channel (grayscale) 3D volume was provided as input to the model and a softmax logit vector with three classes was obtained as output. During the model training process various preprocessing steps were applied to the data. Some of these include:

- -Otsu mask and morphological operations
- -Bounding-box crop with 112^3 patch cropping
- -Z-score normalization

in the form of a modelling process. The aim of these processes was to use the data in the most efficient way and to increase the learning success of the model. In addition, in order to prevent low success due to class imbalances, training was carried out by taking equal number of samples from each class.

The aim of these processes was to utilize the data as efficiently as possible and to increase the learning success of the model. Additionally to prevent low success arising from class imbalances an equal number of samples were taken from each class for training.

After the preprocessing steps various operations were performed on the model. The Inception block contains four parallel branches and different sized filters operate in each branch. While the outputs are combined Squeeze-and-Excitation (SE) is also used to calculate the weight of each channel. In this way the data coming from each layer is computed and information about which data belongs to which class is classified. After all these adjustments fine-tuning was started on the model and during this process 20 epochs and a batch size of 4 were preferred. The AdamW mechanism was used as the optimizer. The tables and graphs obtained at the end of the learning process are presented below:

Table 5.2 Inception Test Results

Metric	Precision	Recall	F1-Score	Support	Accuracy
Weighted Average	0.88	0.88	0.88	90	0.8778

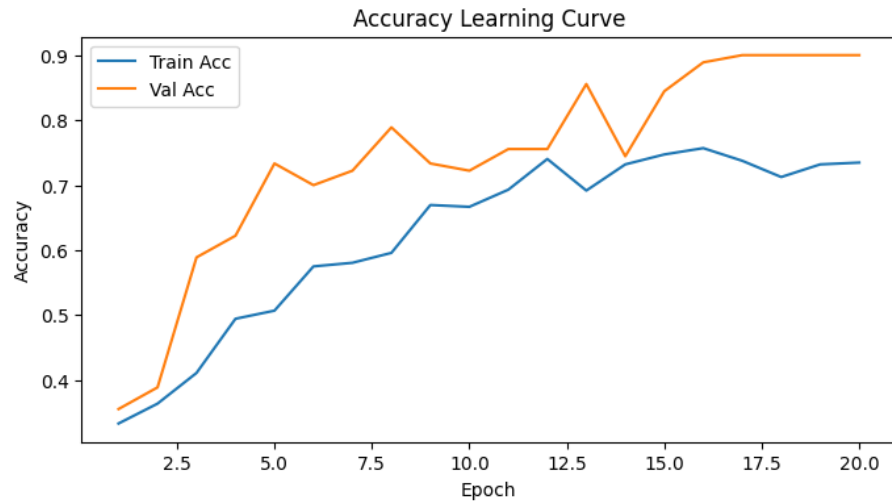


Figure 5.2 Inception Accuracy Graph

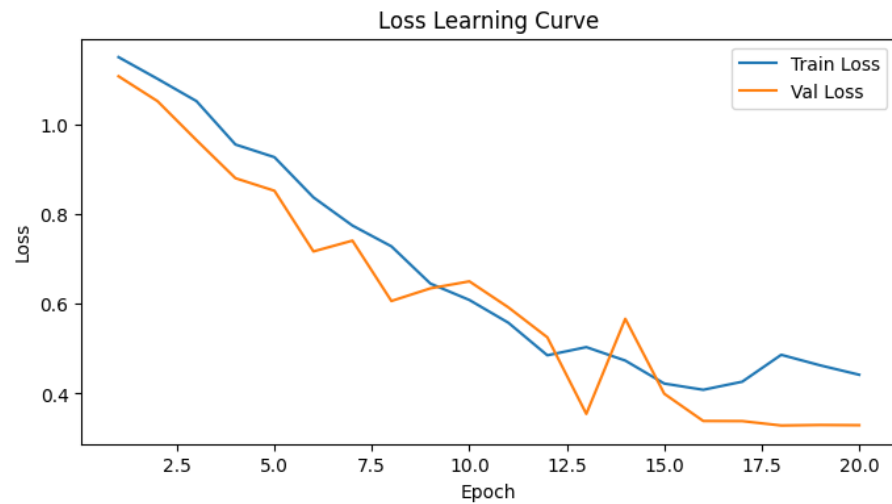


Figure 5.3 Inception Loss Graph

Also same model tested with 5 fold and each fold have 50 epoch to keep better result. Because the lack of the computing units, environment was stopped the model fine-tune process after 12 hours. Just before stopping the fine-tune, we take good results from it. In the 25. epoch process was stopped and below, we add an error message from kaggle envirenment:

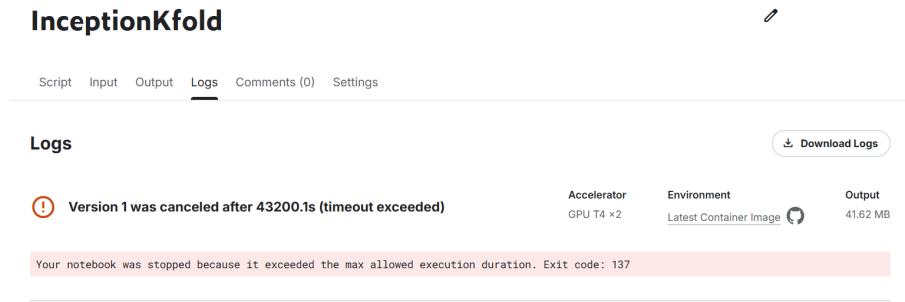


Figure 5.4 Inception 5-folds error message

5.1.1.3 3D-ResNet-18

3D-ResNet is a CNN-based model that takes its basic architectural structure from the ResNet-18 model and includes residual blocks trained with the Kinetics-400 dataset. It contains a total of 4 layers and each layer includes BasicBlock3D. During training

Table 5.3 3D ResNet-18 Layer Structure

Layer	Output Channel	Number of Blocks	First Block Stride
layer1	64	2	(1,1,1)
layer2	128	2	(2,2,2) Spatial+Depth↓
layer3	256	2	(2,2,2)
layer4	512	2	(2,2,2)

20 epochs and a batch size of 2 were preferred. Similarly AdamW was chosen as the optimizer and the input size was given as 64x112x112, while the output was obtained in the size of 64x64x56x56 and the entire volume was reduced to a single vector. To improve the low results arising from class imbalances weight assignments were made inversely proportional to the sample size of each class. The outputs obtained from the training are as follows:

Table 5.4 ResNet Test Results

Metric	Precision	Recall	F1-Score	Support	Accuracy
Weighted Average	0.83	0.83	0.83	219	0.8265

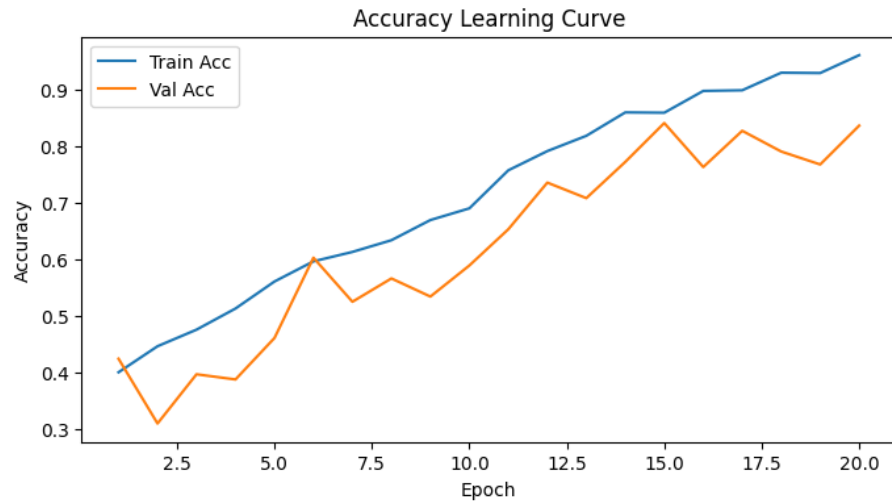


Figure 5.5 3D-ResNet-18 Accuracy Graph

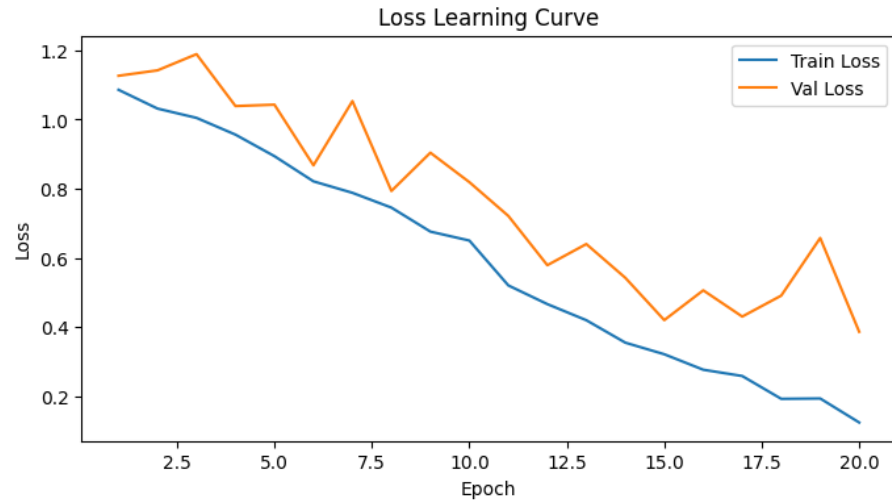


Figure 5.6 3D-ResNet-18 Loss Graph

5.1.1.4 R(2+1)D Classifier

In the model with a ResNet-based architecture each layer is present in a separated manner. This structure is in 2D spatial and 1D temporal form. Thanks to this separation the parameter and computational load decrease allowing for the separate modeling of temporal and spatial features. The input size used was 96x128x128. During the training process 30 epochs were preferred. The reason for choosing a higher number of epochs is that the model is run in pretrained=false mode. Similar to other models a batch size of 2 was preferred and AdamW was also chosen as the optimizer. The training outputs are provided below:

Table 5.5 R2plus1D Test Results

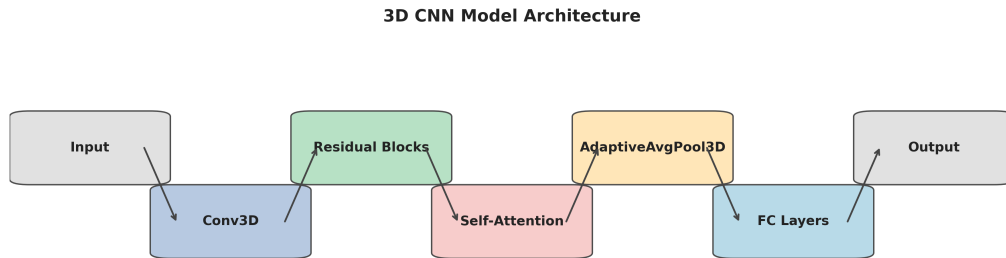
Metric	Precision	Recall	F1-Score	Support	Accuracy
Weighted Average	0.68	0.67	0.66	219	0.6712

5.1.1.5 Custom and Simple CNN Architecture

Another model to be trained with raw data was the model we wrote with an approach based on CNN architecture. We trained this model in 2 fold training cycles with 15 epochs in each fold. Since the results were about 50%, we did not include it in the performance analysis, but we found it appropriate to report this approach

Model architecture:

- Large ($5 \times 5 \times 5 \times 5$) convolution in the first layer + BatchNorm + LeakyReLU + Dropout3d
- Three ResidualBlock ($64 \rightarrow 128 \rightarrow 256 \rightarrow 512$ channels)
- Self-Attention module, followed by AdaptiveAvgPool3d
- A deep classifier: two intermediate fully-connected layers with interlayer LeakyReLU and a high dropout rate

**Figure 5.7** Custom and Simple CNN Model Architecture

5.1.2 Trained Model Outputs With Pre-processed Data

As a result of the training conducted with the data we have the success rate did not reach the desired level so a series of operations were performed. The primary operation among these is the application of data augmentation processes. Various data augmentation techniques were applied to the data to increase the model's generalization ability and reduce overfitting. These processes were applied equally to all classes and the transformations used were implemented with the

TorchIO library. To provide directional diversity by randomly flipping the images along the left-right front-back or top-bottom axes the RandomFlip class was used. Structural variations were employed to increase the model's spatial flexibility and the RandomAffine class was utilized for these operations. Random noise was added using the RandomNoise class to ensure that the model is more robust against low-quality or noisy data and to simulate the magnetic field distortions commonly seen in MRI images the RandomBiasField class was preferred for the model's adaptation to real-world variations. Additionally due to the presence of different numbers of data for each class and the scarcity of data in the AD class the data in the AD class was duplicated while preserving its original form effectively doubling the amount. This process aimed for a more balanced learning during training.

5.1.2.1 Processed Data with Fast Surfer

5.1.2.1.1 orig.mgz data The data obtained from FastSurfer is located in different folders in various formats. When looking at the orig.mgz files created it can be seen that these data are close to the raw data. The data in this file contains the anatomically referenced version of the raw data. The trainings conducted with these data are provided below.

The other folder structure contains the main data expected to be obtained from FastSurfer. These data are located in the folder under the name aparc.DKTatlas+aseg.deep.mgz which is the result of the segmentation and cortical parcellation of the raw data. The trainings conducted in this context are provided below.

DenseNet The DenseNet was fine-tuned with preprocessed data. Considering these files, it is seen that these data are close to the raw data. In the data in this file, there is an anatomically referenced version of the raw data. Additionally, the model was train with 2180 3D MRI images.

Table 5.6 DenseNet Results

Metric	Precision	Recall	F1-Score	Support	Accuracy
Weighted Avg	0.8247	0.7765	0.7785	264	0.7765

Inception3D The inception model, which was previously trained with raw data, was this time trained with preprocessed data. The results are shared below:

Table 5.7 Inception Test Results

Metric	Precision	Recall	F1-Score	Support	Accuracy
Weighted Avg	0.86	0.84	0.84	90	0.8444

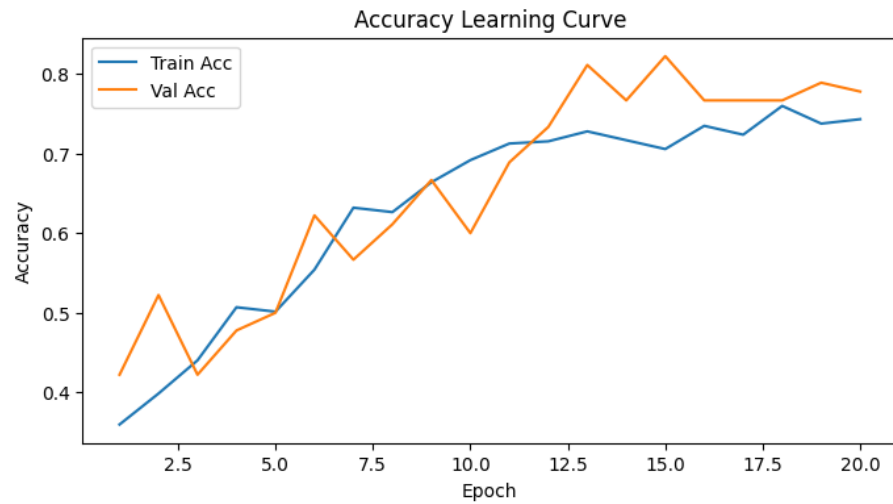


Figure 5.8 Inception Accuracy Graph

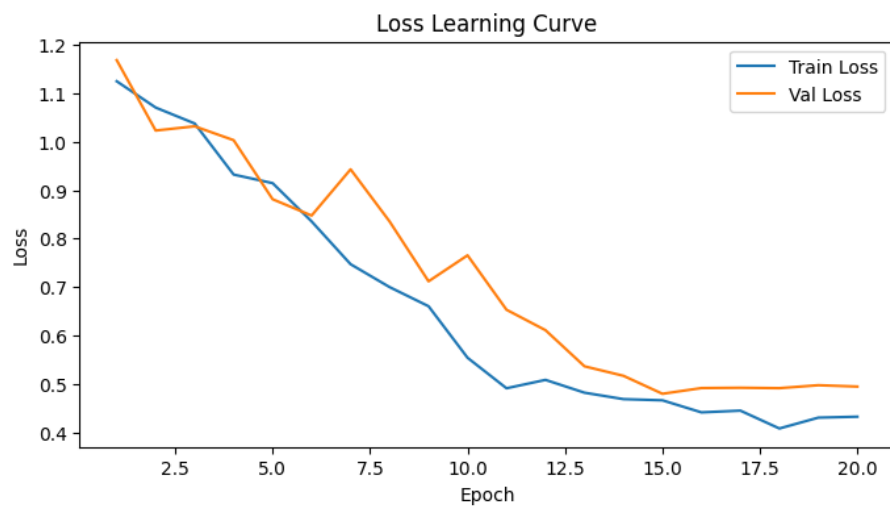


Figure 5.9 Inception Loss Graph

Also, same model fine-tune with 2 fold and each fold have 20 epoch. The comparative results are below:

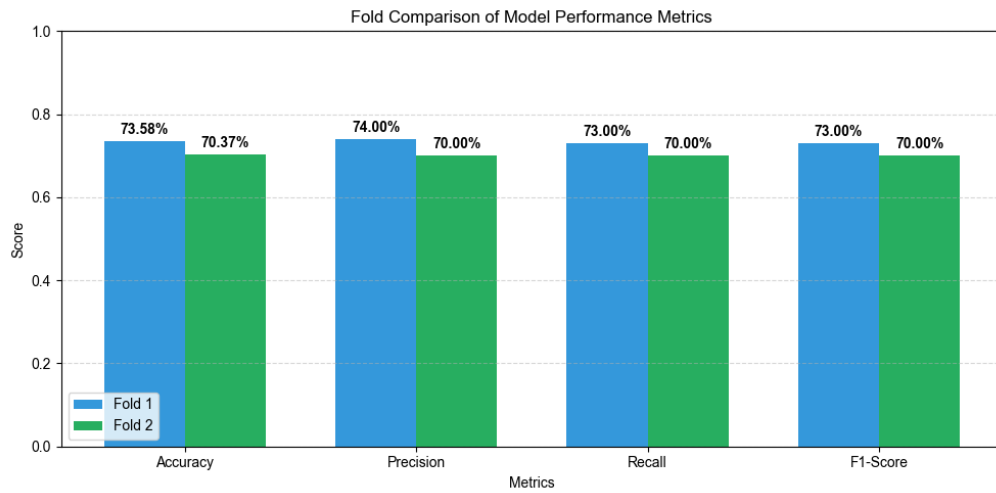


Figure 5.10 Comparison of 2 Folds for Inception

ResNet The outputs of the model trained with the data obtained with the fast surfer process are given below:

Table 5.8 Inception Test Results

Metric	Precision	Recall	F1-Score	Support	Accuracy
Weighted Avg	0.73	0.73	0.73	219	0.7260

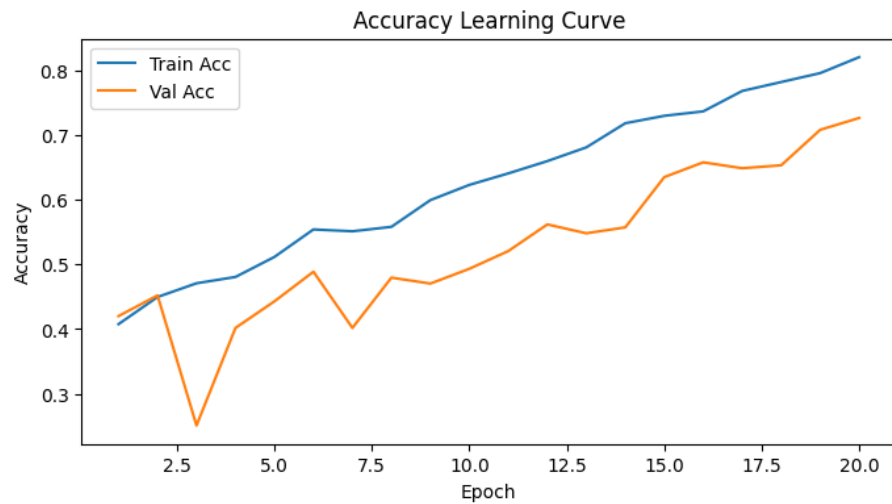


Figure 5.11 3D-ResNet-18 Accuracy Graph

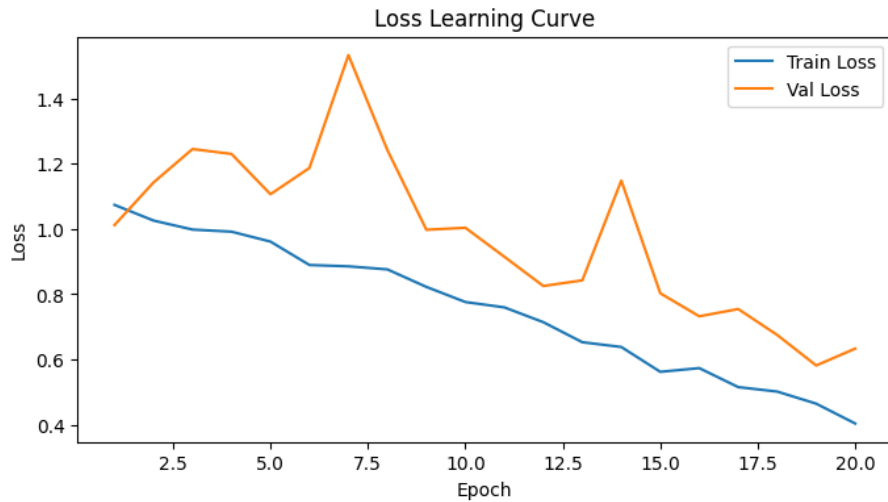


Figure 5.12 3D-ResNet-18 Loss Graph

5.1.2.1.2 aparcmgz The preprocessed data is the result of segmentation and cortical parcellation of raw data.

DenseNet The DenseNet was fine-tuned with aparcmgz files from fastsurfer outputs.

Table 5.9 DenseNet Results

Metric	Precision	Recall	F1-Score	Support	Accuracy
Weighted Avg	0.8414	0.7148	0.6868	263	0.7148

5.1.2.2 Skull-Stripping

This process involves the automatic removal of all structures outside the brain from MRI images such as the skull, skin, adipose tissue, and eyes. The trainings conducted as a result of this preprocessing are provided below.

5.1.2.2.1 Inception Inception model was fine-tuned with preprocessed data. This data just have brain MRI image that the skull and etc removed from MRI images. However, the results are not good because we cannot run it up to 12 hours. We fine-tune with 5 folds and each folds have 10 epoch but the process was interrupted on the middle of the 3.fold. Therefore, we cannot interpret correctly the result.

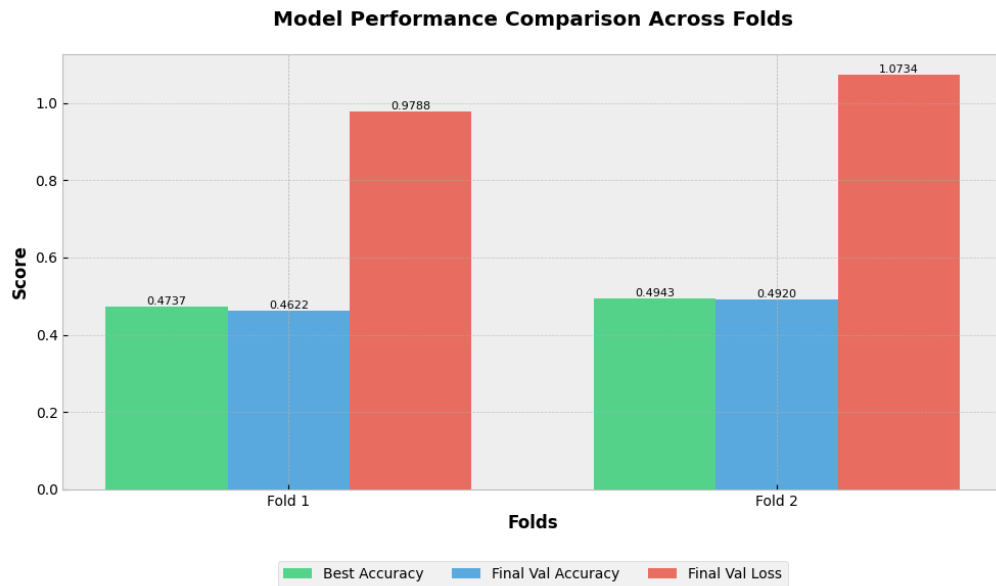


Figure 5.13 Inception Test Results with Skull-Stripped Data

5.1.2.2.2 DenseNet Training was performed with a dataset in which all structures other than the brain such as skull, skin, adipose tissue and eyes were automatically removed from MRI images.

Table 5.10 DenseNet Results

Metric	Precision	Recall	F1-Score	Support	Accuracy
Weighted Avg	0.6745	0.5057	0.4741	263	0.5057

5.2 Comparative Analysis

5.2.1 Same Dataset With Different Models

There was fine-tuned 4 different model with raw data. The comparative tables are below:

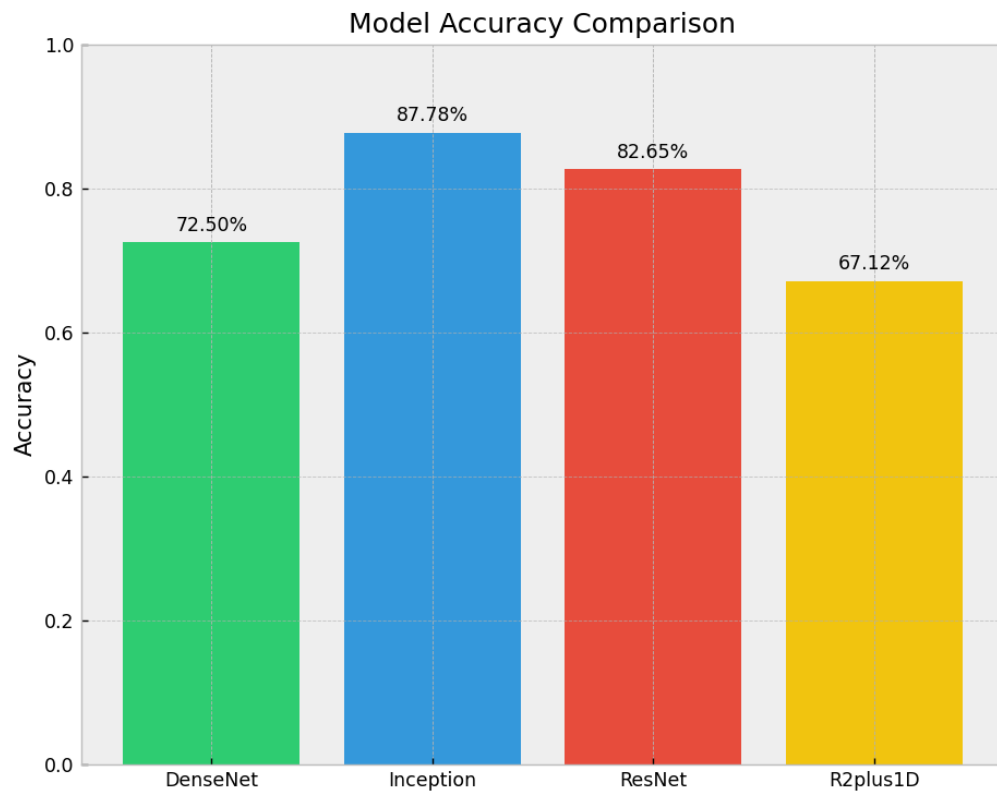


Figure 5.14 Accuracy Values From 4 Different Models

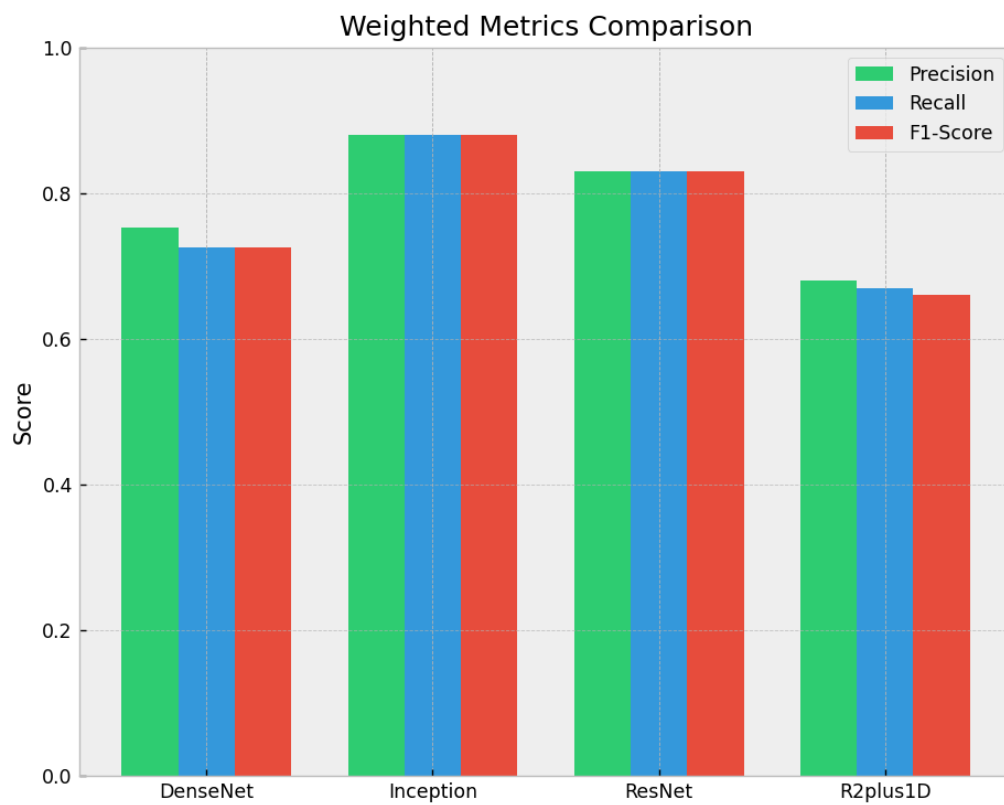


Figure 5.15 Different Metrics From 4 Models

There are also 3 different models, one of this with 2 different combination, fine-tuned with preprocess data. The data preprocessed with fastsurfer tool. The comparative tables are below:

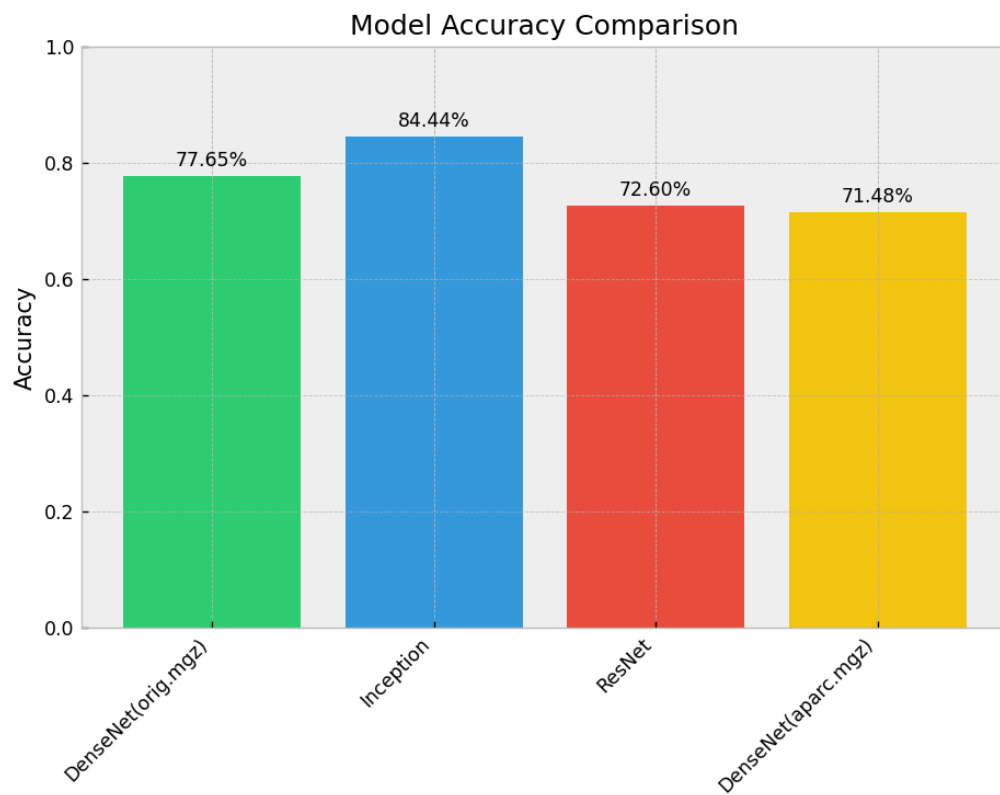


Figure 5.16 Accuracy Values From 4 Different Models

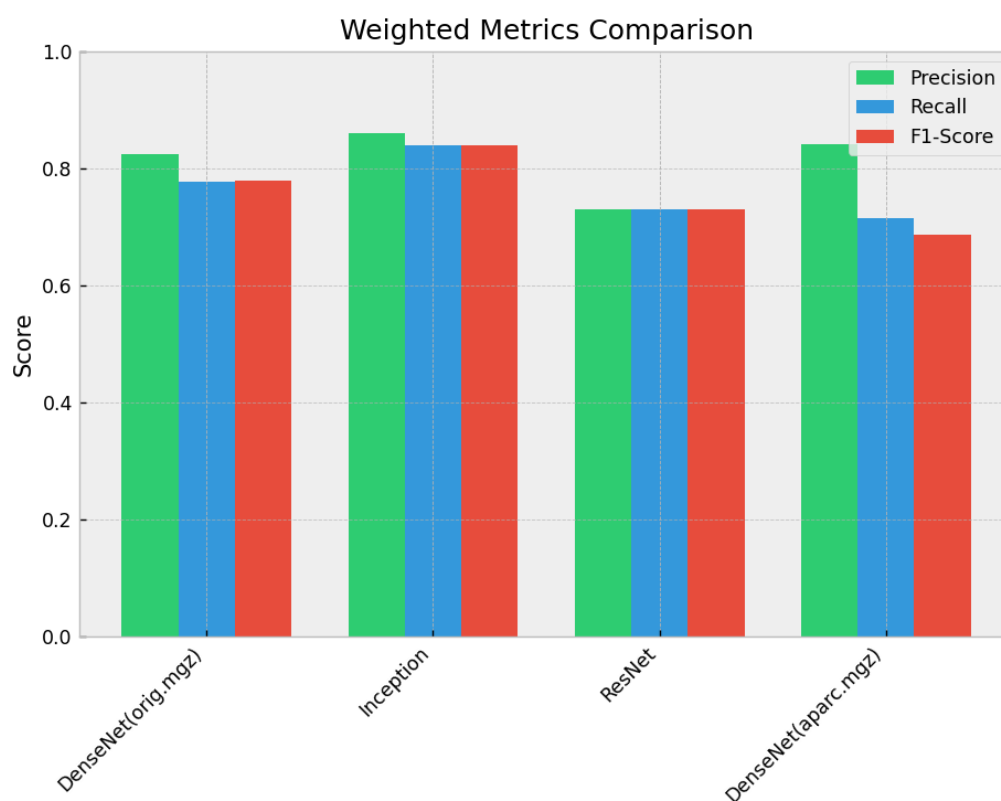


Figure 5.17 Different Metrics From 4 Models

5.2.2 Same Models on Different Dataset

The tables comparing the success of the same models according to the results obtained from the experimental results are given below:

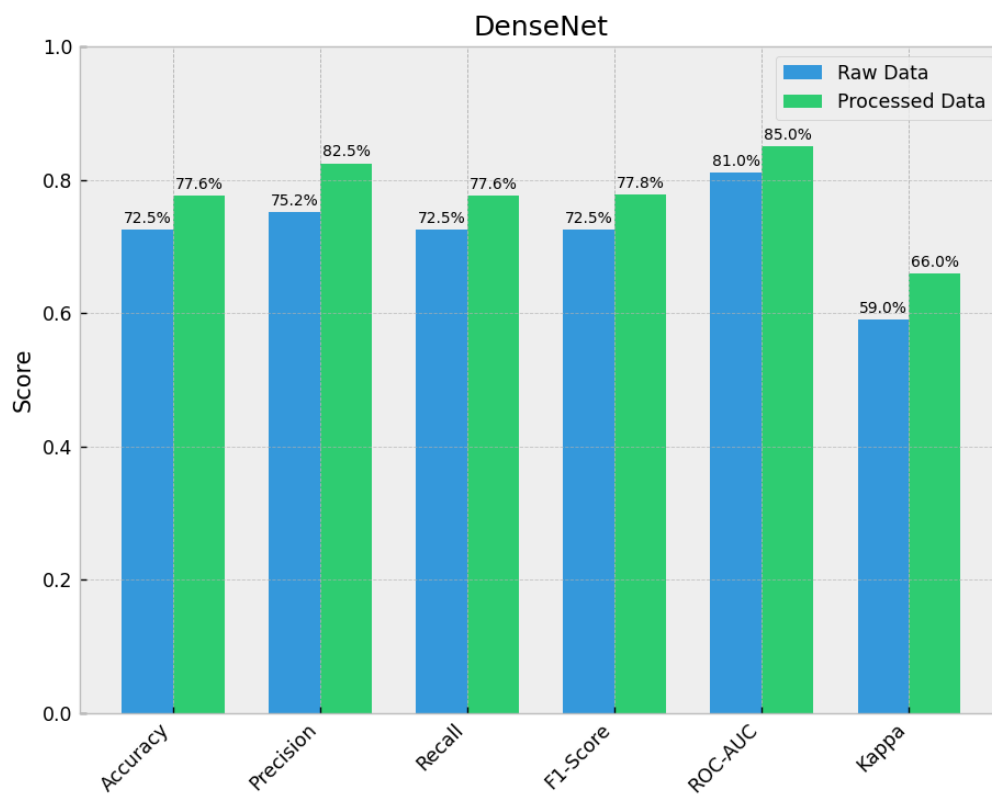


Figure 5.18 DenseNet Performance Metrics

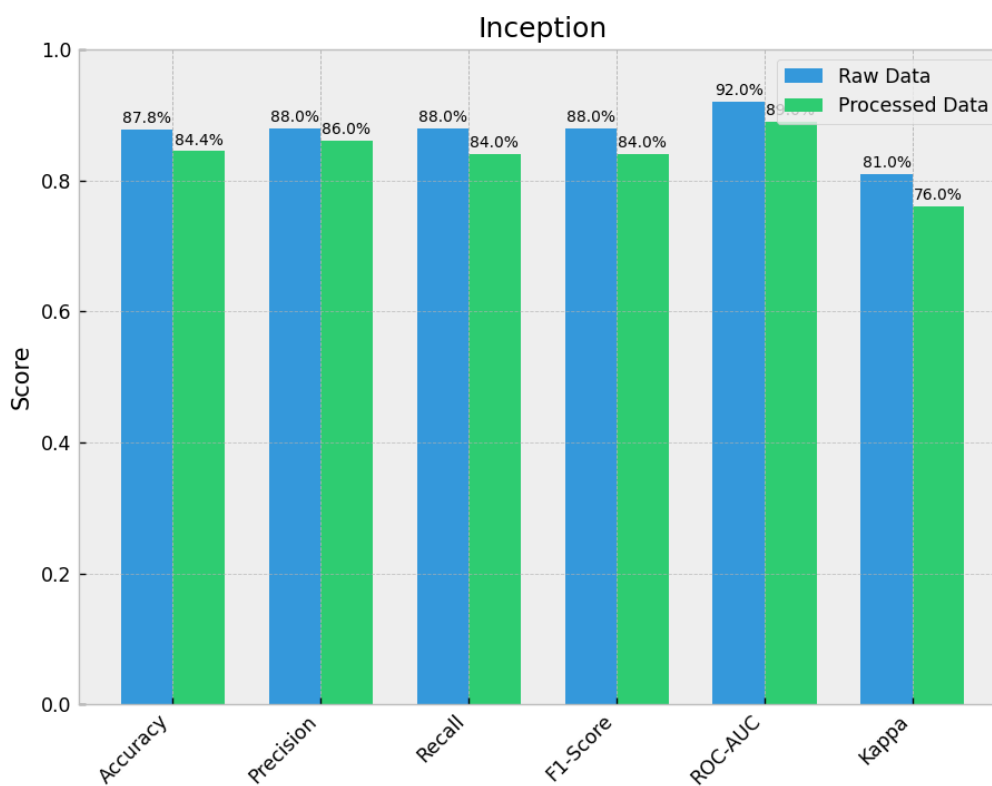


Figure 5.19 Inception Performance Metrics

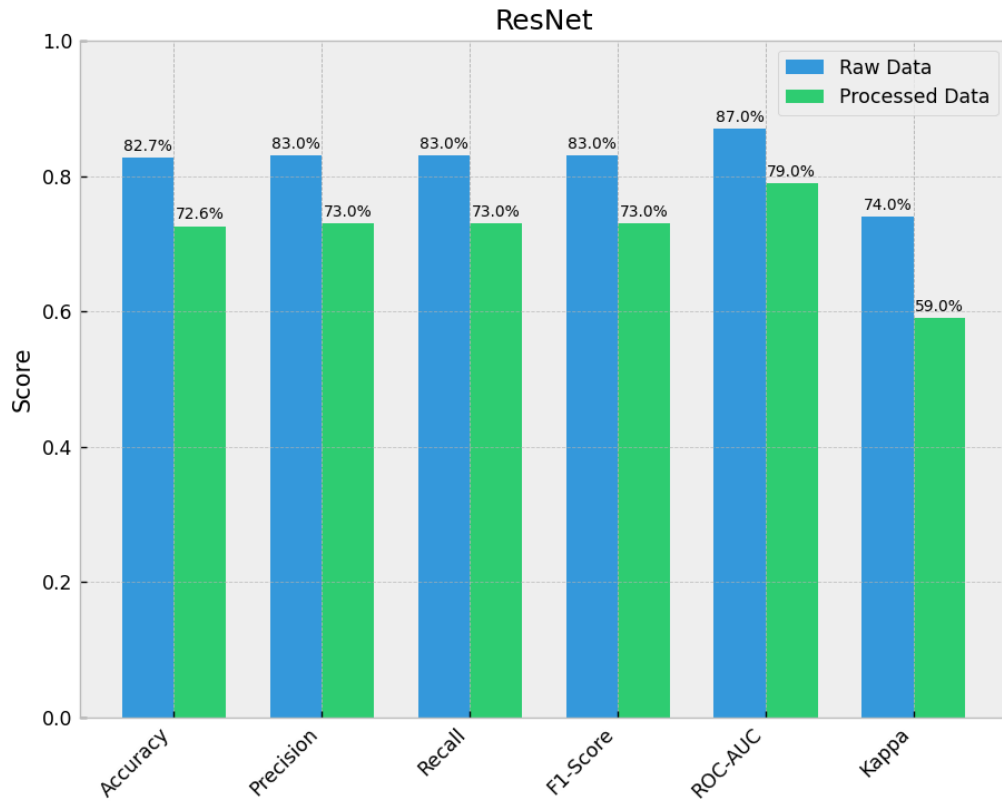


Figure 5.20 ResNet Performance Metrics

5.2.3 Model Performances on Machine Learning Algorithms

Inception, DenseNet and ResNet models were subjected to 8 different machine learning algorithms(Random Forest, Gradient Boosting, SVM, KNN, Logistic Regression, Naive Bayes, Decision Tree, Neural Network). The main purpose here was to confirm the accuracy of the parameters obtained during the training process and to test the model performances with algorithms designed for different purposes.

Table 5.11 Performance Metrics of Machine Learning Models with Different CNN Architectures

ML Models CNN Models	ROC-AUC			F1 Score			Kappa		
	DenseNet	Inception	ResNet	DenseNet	Inception	ResNet	DenseNet	Inception	ResNet
Decision Tree	0.5330	0.4820	0.5210	0.5322	0.4878	0.5213	0.5310	0.4800	0.5200
Gradient Boosting	0.6760	0.5860	0.6880	0.6746	0.5809	0.6859	0.6740	0.5840	0.6850
KNN	0.6270	0.5390	0.6510	0.6255	0.5338	0.6475	0.6250	0.5370	0.6470
Logistic Regression	0.6130	0.4900	0.6210	0.6081	0.4767	0.6201	0.6110	0.4880	0.6180
Naive Bayes	0.4070	0.3640	0.3720	0.4006	0.3360	0.3358	0.4050	0.3620	0.3350
Neural Network	0.5990	0.5090	0.6540	0.5979	0.4744	0.6496	0.5970	0.5070	0.6480
Random Forest	0.6460	0.5620	0.6450	0.6422	0.5579	0.6418	0.6430	0.5600	0.6420
SVM	0.5950	0.4180	0.6320	0.5813	0.3870	0.6294	0.5920	0.4150	0.6280

Table 5.12 Performance Metrics of Machine Learning Models with Different CNN Architectures

ML Models CNN Models	ROC-AUC			F1 Score			Kappa		
	DenseNet	Inception	ResNet	DenseNet	Inception	ResNet	DenseNet	Inception	ResNet
Decision Tree	0.5330	0.4820	0.5210	0.5322	0.4878	0.5213	0.5310	0.4800	0.5200
Gradient Boosting	0.6760	0.5860	0.6880	0.6746	0.5809	0.6859	0.6740	0.5840	0.6850
KNN	0.6270	0.5390	0.6510	0.6255	0.5338	0.6475	0.6250	0.5370	0.6470
Logistic Regression	0.6130	0.4900	0.6210	0.6081	0.4767	0.6201	0.6110	0.4880	0.6180
Naive Bayes	0.4070	0.3640	0.3720	0.4006	0.3360	0.3358	0.4050	0.3620	0.3350
Neural Network	0.5990	0.5090	0.6540	0.5979	0.4744	0.6496	0.5970	0.5070	0.6480
Random Forest	0.6460	0.5620	0.6450	0.6422	0.5579	0.6418	0.6430	0.5600	0.6420
SVM	0.5950	0.4180	0.6320	0.5813	0.3870	0.6294	0.5920	0.4150	0.6280

In this study the Alzheimer Disease Neuroimaging Initiative (ADNI) dataset was used for the development and evaluation of deep learning models related to Alzheimer’s disease (AD). This dataset consists of participants aged between 55 and 90 and is divided into three groups: cognitively normal individuals (CN), individuals with mild cognitive impairment (MCI) and individuals diagnosed with Alzheimer’s disease (AD). The primary aim of the ADNI dataset is to focus on the development of biomarkers as outcome measures for clinical trials and to facilitate the investigation of biomarkers in the early stages of the disease. In this context the dataset has become an important resource in our research on Alzheimer’s disease.

6.1 Dataset

A significant feature of the ADNI dataset is its inclusion of various data types. In the neuroimaging category different types of data such as structural magnetic resonance imaging (MRI), positron emission tomography (PET) and diffusion tensor imaging (DTI) are available. Additionally biological data such as cerebrospinal fluid and blood biomarkers are also included in the dataset. This diversity has provided us with a rich foundation during the training and evaluation phases of deep learning models.

6.2 Preprocessing Phase

The preprocessing phase conducted using FastSurfer has played a critical role in enhancing data quality and ensuring model accuracy in our study. Brain segmentation was performed through deep learning-focused anatomical segmentation which replicated the outputs of traditional FreeSurfer pipelines in a much shorter time. The alignment of the segmented data with FreeSurfer facilitated the integration of surface extraction processes. This phase allowed for the calculation of morphometric measurements such as cortical thickness, curvature and surface area. Furthermore all MRI volumes underwent systematic visual inspection taking into account criteria such

as brain mask alignment absence of extra-cerebral tissue artifacts, and anatomical accuracy. This detailed quality control process has increased the reliability of the obtained data and positively influenced the success of the model.

6.3 Model Development and Fine-Tuning

In this study deep learning architectures such as DenseNet and ResNet were utilized for fine-tuning on ADNI MRI volumes with pre-trained weights. Specifically two different fine-tuning strategies (feature extractor mode and gradual unfreezing mode) were evaluated. In feature extractor mode all convolutional layers were frozen and only the newly added classification heads were trained. This approach allowed for the preservation of general low-level features while adapting to Alzheimer-specific patterns. In gradual unfreezing mode deeper convolutional blocks were sequentially unfrozen and fine-tuned. This strategy has effectively facilitated the refinement of domain-specific mid and high-level features.

The results obtained were supported by optimized hyperparameters aimed at enhancing the model's accuracy. The DenseNet-121 variant supports feature reuse due to its dense connectivity structure and reduces the vanishing gradient problem often seen in deep architectures. ResNet-50 on the other hand, facilitates the training of very deep networks thanks to its residual shortcut connections. The use of these architectures has contributed to achieving significant success in the classification of Alzheimer's disease.

The obtained results have been presented in detail to compare the performance of different models. For example, the DenseNet model exhibited remarkable performance with a precision of 84.14% and an accuracy of 71.48%. On the other hand, the Inception model yielded lower results based on the characteristics of the dataset. Particularly in the fine-tuning processes with preprocessed data, the interruption of the model's training process made it impossible to accurately interpret the results. This situation stands out as a factor that directly affects the model's performance.

6.4 Comparative Analysis

The fine-tuning processes of different models on the same dataset play an important role in model selection. The findings indicate that deep learning architectures can be effective in the diagnosis and classification of Alzheimer's disease. However the quality of the dataset and the training processes of the model directly affect the reliability of

the obtained results. For instance the fine-tuning processes with DenseNet enabled higher accuracy rates while the limited performance of the Inception model stemmed from issues related to the model's complexity and the management of the training process.

6.5 Future Work

In conclusion this study has highlighted the potential of deep learning applications in Alzheimer's disease and emphasized the importance of dataset quality and model architecture. Future research could deepen knowledge in this area by utilizing more comprehensive datasets and more complex model architectures. Additionally testing the findings with real-world data and enhancing their validity is an important detail for integrating the model into clinical applications. In this regard adopting a multidisciplinary approach could accelerate progress in both clinical and research fields. This study has demonstrated the applicability of deep learning techniques in Alzheimer's disease research and has laid a foundation for future studies.

References

- [1] W. H. Organization. "Dementia fact sheet." (2022), [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/dementia> (visited on 03/27/2025).
- [2] M. M. Mielke, P. Vemuri, and W. A. Rocca, "Clinical epidemiology of alzheimer's disease: Assessing sex and gender differences," *Clinical Epidemiology*, vol. 6, no. 1, pp. 37–48, 2014.
- [3] G. McKhann, D. S. Knopman, H. Chertkow, *et al.*, "The diagnosis of dementia due to alzheimer's disease: Recommendations from the national institute on aging–alzheimer's association workgroups on diagnostic criteria," *Alzheimer's & Dementia*, vol. 7, no. 3, pp. 263–269, 2011.
- [4] R. S. Doody, R. G. Thomas, M. Farlow, *et al.*, "Guidelines for drug trials in the treatment of age-associated cognitive decline and alzheimer's disease," *Alzheimer's Research & Therapy*, 2008, Details to be updated.
- [5] S. L. Mitchell, J. M. Teno, D. K. Kiely, *et al.*, "The clinical course of advanced dementia," *New England Journal of Medicine*, vol. 361, no. 16, pp. 1529–1538, 2009.
- [6] C. R. Jack, D. A. Bennett, K. Blennow, *et al.*, "Nia-aa research framework: Toward a biological definition of alzheimer's disease," *Alzheimer's & Dementia*, vol. 14, no. 4, pp. 535–562, 2018.
- [7] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "Mini-mental state: A practical method for grading the cognitive state of patients for the clinician," *Journal of Psychiatric Research*, vol. 12, pp. 189–198, 1975.
- [8] A. Khvostikov, K. Aderghal, J. Benois-Pineau, A. Krylov, and G. Catheline, *3d cnn-based classification using smri and md-dti images for alzheimer disease studies*, arXiv preprint: 1801.05968, available at <https://arxiv.org/abs/1801.05968>, 2018.
- [9] R. Sampath and M. Baskar, "Alzheimer's disease prediction using fly-optimized densely connected convolution neural networks based on mri images," 2024, Published in *The Journal of Prevention of Alzheimer's Disease*, Volume 11, Issue 4, Pages 1106-1121. DOI: <https://doi.org/10.14283/jpad.2024.66>, Available at <https://www.sciencedirect.com/science/article/pii/S2274580724000372>.
- [10] A. Ebrahimi, S. Luo, and R. Chiong, "Deep sequence modelling for alzheimer's disease detection using mri," 2021, Published in *Computers in Biology and Medicine*, Volume 134, Article 104537. DOI: <https://doi.org/10.1016/j.combiomed.2021.104537>, Available at <https://www.sciencedirect.com/science/article/pii/S0010482521003310>.

- [11] R. V. Marinescu *et al.*, “The alzheimer’s disease prediction of longitudinal evolution (tadpole) challenge: Results after 1 year follow-up,” *arXiv preprint*: 2002.03419, available at <https://arxiv.org/abs/2002.03419>, 2021.
- [12] R. V. Marinescu *et al.*, “Tadpole challenge: Accurate alzheimer’s disease prediction through crowdsourced forecasting of future data,” Presented in **Predictive Intelligence in Medicine**, Springer International Publishing. Available at <https://tadpole.grand-challenge.org/>, 2019.
- [13] L. Henschel, Y. O. Halchenko, S. Tak, *et al.*, “Fastsurfer – a fast and accurate deep learning based neuroimaging pipeline,” *Neuroinformatics*, vol. 18, no. 2, pp. 265–281, 2020.
- [14] A. Loddo, S. Buttau, and C. D. Ruberto, “Deep learning based pipelines for alzheimer’s disease diagnosis: A comparative study and a novel deep-ensemble method,” 2022, Published in **Computers in Biology and Medicine**. Available at <https://www.sciencedirect.com/science/article/pii/S001048252100826X>.
- [15] V. Sathiyamoorthi, A. Ilavarasi, K. Murugeswari, S. T. Ahmed, B. A. Devi, and M. Kalipindi, “A deep convolutional neural network based computer aided diagnosis system for the prediction of alzheimer’s disease in mri images,” *Measurement*, 2021, Computer-aided diagnosis system using deep CNN.
- [16] K. Kruthika, Rajeswari, and H. Maheshappa, “Cbir system using capsule networks and 3d cnn for alzheimer’s disease diagnosis,” *Informatics in Medicine Unlocked*, 2019, Computer-aided diagnosis system using Capsule Networks and 3D CNN.
- [17] R. Jain, N. Jain, A. Aggarwal, and D. J. Hemanth, “Convolutional neural network based alzheimer’s disease classification from magnetic resonance brain images,” *Cognitive Systems Research*, 2019, Computer-aided diagnosis system using CNN for Alzheimer’s classification.
- [18] M. A. Ben, Y. Taoufik, and R. A. Ben, “Mri images analysis method for early stage alzheimer’s disease detection,” *International Journal of Computer Science and Network Security*, 2020, Early-stage Alzheimer’s detection using MRI image analysis.
- [19] S. Foroughipoor, K. Moradi, and H. Bolhasani, “Alzheimer’s disease diagnosis by deep learning using mri-based approaches,” *arXiv preprint*, 2023, Deep learning approaches for Alzheimer’s diagnosis using MRI. [Online]. Available: <https://arxiv.org/abs/2310.17755>.
- [20] S. Li, H. Qu, X. Dong, B. Dang, H. Zang, and Y. Gong, “Leveraging deep learning and xception architecture for high-accuracy mri classification in alzheimer diagnosis,” *arXiv preprint*, 2024, Deep learning and Xception architecture for Alzheimer’s MRI classification. [Online]. Available: <https://arxiv.org/abs/2403.16212>.

- [21] S. Basaia *et al.*, “Automated classification of alzheimer’s disease and mild cognitive impairment using a single mri and deep neural networks,” *NeuroImage: Clinical*, vol. 21, p. 101645, 2019, Deep learning using CNN for Alzheimer’s disease and MCI classification from MRI. DOI: <https://doi.org/10.1016/j.nicl.2018.101645>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2213158218303930>.
- [22] S. Shojaei, M. Saniee Abadeh, and Z. Momeni, “An evolutionary explainable deep learning approach for alzheimer’s mri classification,” *Expert Systems with Applications*, vol. 220, p. 119709, 2023, Evolutionary explainable deep learning for Alzheimer’s classification using MRI and CNN. DOI: <https://doi.org/10.1016/j.eswa.2023.119709>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417423002105>.
- [23] J. Rasheed, M. U. Shaikh, M. Jafri, A. U. Khan, M. Sandhu, and H. Shin, “Leveraging capsnet for enhanced classification of 3d mri images for alzheimer’s diagnosis,” *Biomedical Signal Processing and Control*, vol. 103, p. 107384, 2025, Enhanced Alzheimer’s diagnosis using Hybrid CapsNet-3D model with hippocampal volume estimation. DOI: <https://doi.org/10.1016/j.bspc.2024.107384>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1746809424014423>.
- [24] S. Venkat, T. Ghodeswar, P. Chavan, S. K. Narayanasamy, and K. Srinivasan, “Mri-based automated diagnosis of alzheimer’s disease using alzh-net deep learning model,” *Biomedical Signal Processing and Control*, vol. 102, p. 107367, 2025, Automated Alzheimer’s diagnosis using Alzh-Net deep learning model based on MRI data. DOI: <https://doi.org/10.1016/j.bspc.2024.107367>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1746809424014253>.
- [25] E. U. Haq, Q. Yong, Z. Yuan, X. Huarong, and R. U. Haq, “Multimodal fusion diagnosis of alzheimer’s disease via lightweight cnn-lstm model using magnetic resonance imaging (mri),” *Biomedical Signal Processing and Control*, vol. 104, p. 107545, 2025, Multimodal fusion using CNN-LSTM for Alzheimer’s diagnosis based on MRI data. DOI: <https://doi.org/10.1016/j.bspc.2025.107545>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1746809425000564>.
- [26] Alzheimer’s Disease Neuroimaging Initiative, *Alzheimer’s disease neuroimaging initiative (adni)*, <https://adni.loni.usc.edu/>, Accessed: 2025-05-01, 2025.
- [27] Laboratory of Neuro Imaging (LONI), USC Mark and Mary Stevens Neuroimaging and Informatics Institute, *Image & Data Archive (IDA) Login Page*, <https://ida.loni.usc.edu/login.jsp>, Accessed: 2025-05-01, 2025.
- [28] Deep-MI and FastSurfer Contributors, *FastSurfer – A Fast and Accurate Deep Learning Based Neuroimaging Pipeline (GitHub Repository)*, <https://github.com/Deep-MI/FastSurfer>, GitHub repository, accessed: 2025-05-01, 2025.
- [29] H. E. Kim, A. Cosa-Linan, N. Santhanam, and et al., “Transfer learning for medical image classification: A literature review,” *BMC Medical Imaging*, vol. 22, p. 69, 2022. DOI: 10.1186/s12880-022-00793-7. [Online]. Available: <https://doi.org/10.1186/s12880-022-00793-7>.

- [30] A. Davila, J. Colan, and Y. Hasegawa, "Comparison of fine-tuning strategies for transfer learning in medical image classification," *arXiv preprint arXiv:2406.10050v1*, 2024, Accessed: 2025-05-01. [Online]. Available: <https://arxiv.org/abs/2406.10050v1>.
- [31] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, *Densely connected convolutional networks*, Accessed: 2025-05-01, 2018. arXiv: 1608 . 06993 [cs.CV]. [Online]. Available: <https://arxiv.org/abs/1608.06993>.
- [32] C. Zhang *et al.*, "Resnet or densenet? introducing dense shortcuts to resnet," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Accessed: 2025-05-01, 2021, pp. 3550–3559. DOI: 10 . 1109 / WACV48630 . 2021 . 00362. [Online]. Available: https://openaccess.thecvf.com/content/WACV2021/papers/Zhang_ResNet_or_DenseNet_Introducing_Dense_Shortcuts_to_ResNet_WACV_2021_paper.pdf.
- [33] A. W. Saleh, G. Gupta, S. B. Khan, N. A. Alkhaldi, and A. Verma, "An alzheimer's disease classification model using transfer learning densenet with embedded healthcare decision support system," *Decision Analytics Journal*, vol. 9, p. 100348, Dec. 2023, Accepted: October 25, 2023; Online: November 8, 2023; License: CC BY 4.0; Accessed: 2025-05-01, ISSN: 2772-6622. DOI: 10.1016/j.dajour.2023.100348. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2772662223001881>.

Curriculum Vitae

FIRST MEMBER

Name-Surname: LOGHMAN ORUJOV

Birthdate and Place of Birth: 24.02.2000, Nahçıvan

E-mail: lokmanorucov34@gmail.com

Phone: 0553 715 96 94

Practical Training: KKA software and development

SECOND MEMBER

Name-Surname: ŞEHMUS YAKUT

Birthdate and Place of Birth: 08.03.2002, BATMAN

E-mail: sehmysyakut1@gmail.com

Phone: 0545 386 6209

Practical Training: Turkcell

Project System Informations

System and Software: Windows Operating System, Python

Required RAM: 16GB

Required Disk: 100GB