AN AI BASED ALZHEIMER'S CLASSIFICATION SYSTEM USING MRI IMAGES AND XAI

BY

AZIE CHIMDIUME

GOU/U23/CSC/1315

DEPARTMENT OF COMPUTER SCIENCE, FACULTY OF COMPUTING AND INFORMATION TECHNOLOGY

GODFREY OKOYE UNIVERSITY, ENUGU STATE.

IN PARTIAL FULFILLMENT OF THE AWARD OF BACHELOR OF

SCIENCE (B.SC), DEGREE IN COMPUTER SCIENCE.

SUPERVISOR: ENGR DR.UGORGI CALLISTUS

MAY, 2026

APPROVAL PAGE

This research, titled "AN AI BASED ALZHEIMER'S CLASSIFICATION SYSTEM USING MRI IMAGES AND XAI," has been assessed and approved by the Department of Computer Science and Mathematics Committees in the Faculty of Computer and Information Technology, Godfrey Okoye University, Enugu, Nigeria.

Supervisor	....	....
Engr Dr.Ugorgi Callistus	Signature	Date
External Examiner	....	....
	Signature	Date
Dr. Frank Okebanama	....	....
HOD, Department of Computer Science	Signature	Date
Prof. Ajah Ifeyinwa	....	....
Dean, Faculty of Computing And Information Technology.	Signature	Date

CERTIFICATION

I certify that I completed the research included therein and that it was primarily based on my observations and investigation. Consequently, I will fully accept the University's decision if I am found to have cheated on the assignment.

____

AZIE CHIMDIUME

GOU/U23/CSC/1315

DEDICATION

This project is dedicated to my beloved parents, whose endless love, sacrifice, and unwavering support have been my greatest source of strength throughout this academic journey, and to my supervisor, Engr Dr.Ugorgi Callistus, whose guidance and encouragement made this work possible.

ACKNOWLEDGMENTS

In the first place, I dedicate all glory to Almighty God for His help in making this study a success. I will like to thank my supervisor, Engr Dr.Ugorgi Callistus, for his untold assistance, patience, and dedication towards the successful completion of this research. It must be noted that his guidance, assistance in ensuring that the system works as required and dedication to perfection in documentation were essential ingredients of this work.

Secondly, I would like to specifically thank Engr. Camillus, who has played a very significant role in shaping my academic studies. He assisted in guiding me practically on how to execute the work well and document everything well.

Also, I would like to thank the Head of Department and other lecturers in the Department of Computer Science, Godfrey Okoye University, for the knowledge and guidance provided during my academic studies.

Last but not least, I would like to thank my parents for their love, financial assistance, and encouragement which kept me going.

ABSTRACT

Alzheimer's disease is the major causative factor of dementia in the world with more than 55 million people suffering from it, and the numbers have been increasing drastically in developing nations such as Nigeria, where access to diagnostic services is severely lacking. The early diagnosis of Alzheimer's disease is paramount in ensuring the treatment of the condition through clinical interventions; however, the current method of diagnosing Alzheimer's is very subjective as it uses brain MRI scans interpreted by radiologists, a resource which is in critically short supply in hospitals within low-income settings. The current research aims to provide an end-to-end artificial intelligence based Alzheimer's disease detection and hospital management solution which combines deep learning classifiers and explainable artificial intelligence techniques. The proposed model uses the EfficientNet-B0 Convolutional Neural Network architecture, and its training was achieved through transfer learning on the OASIS Alzheimer's MRI data set consisting of over 90,000 brain MRI images of four different disease stages, namely, NonDemented, VeryMildDemented, MildDemented, and ModerateDemented. Training was done using class weighting to solve the issue of extreme dataset imbalance, while Gradient-weighted Class Activation Mapping (Grad-CAM) was employed as an after-the-fact explainability approach to create heat maps showing the regions of the brain which affected predictions the most. The validation accuracy obtained during training reached 98.66% in the 19th epoch out of a total of 25 training epochs, with the weighted average F1-score of the model reaching 0.8126 on the unseen test data set, consistently emphasizing clinically important parts of the brain such as the hippocampus and ventricles through its Grad-CAM. The trained model was used in a full-stack hospital management web application developed using Flask and vanilla JavaScript, with functionality such as role-based access control for Admin and Doctor roles, patient registration, uploading of MRIs, storing the results, tracking the history of scans and even transferring of patients, all of which were then hosted on a cloud hosting platform. This makes the model clinically deployable since it can be used in hospitals even with regular computational hardware, an important contribution towards early diagnoses of AD in limited resources areas.

TABLE OF CONTENT

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

Alzheimer's disease is an illness that is marked by the progressive loss of memory, cognitive abilities and autonomy. As per the doctors, it is the main contributor to dementia. Brain cells die in this condition and result in brain cell deterioration of specific regions such as hippocampus that helps in memory formation (Forshing Lui; Jack W. Tsao.2024).

Currently, there are more than 55 million dementia patients around the globe, and approximately 60 to 70 percent of them are suffering from Alzheimer's disease. This figure could go up to 139 million in 2050 if nothing changes. Dementia prevalence is increasing in Africa including Nigeria with their aging population. Patients in Africa are still suffering owing to the late diagnosis because by the time it is detected, it is already too late. That is why the significance of early diagnosis can never be overemphasized (WHO, 2023).

Magnetic Resonance Imaging (MRI) is one of the most significant non-invasive technologies that are useful for brain imaging and helps to identify features like cerebral atrophy and enlargement of ventricles. However, the analysis of MRI requires expertise and takes time; therefore, under pressure conditions, minor changes might remain unnoticed even by professionals (National Institute on Aging, 2022).

In the past decade, the evolution of AI, especially Deep Learning, transformed the field of medical imaging. Deep CNN models can analyze thousands of MRIs and detect abnormalities even the most qualified physicians would sometimes miss. There have been some demonstrations that efficient CNNs such as EfficientNet can classify the progression of Alzheimer's disease reliably yet efficient enough to be performed on personal computers (Tan & Le, 2019).

However, many intelligent algorithms have a problem of transparency – they provide output data without any explanations why. For example, a physician should get more information why the AI system classified his patient's disease as Moderate Demented, rather than receiving a number only. Thus, the emergence of Explainable Artificial Intelligence (XAI), such as Grad-CAM, became relevant. Grad-CAM creates color maps over brain images highlighting what parts were considered by the neural network (Selvaraju et al., 2017).

In recent times, such a combination proves to be quite effective. For instance, when applying EfficientNet for large datasets with MRI images, a great accuracy rate was obtained; however, by adding Grad-CAM, medical practitioners were able to better interpret their results. This would greatly benefit the Nigerian healthcare since, sometimes, it becomes challenging for the population in remote rural areas to receive the assistance of an advanced imaging machine or a specialist. Thus, with an efficient AI program installed on a normal computer of a hospital, the situation may change dramatically.

The proposed work develops such a technology. Namely, a system with an AI program is designed. The input to the system is an MRI image which then is classified into one out of four classes: NonDemented, VeryMildDemented, MildDemented, and ModerateDemented. In addition, an explanation in the form of images for the classification results is provided. A simple hospital-based workflow of patient registration and management is implemented.

1.2 Statement of the Problem

Although Alzheimer's is one of the most damaging forms of neurodegenerative diseases in the world, diagnosing the disease early enough is still a major challenge for clinicians. The current procedures for diagnosing Alzheimer's involve the manual reading of MRIs by specialist doctors, which is a time-consuming and subjective process, and is more vulnerable to errors by humans, especially in limited-resource health facilities such as in Nigeria, where there are few specialists. Although there have been breakthroughs in deep learning algorithms for automated MRI-based classification of Alzheimer's, several problems still persist. These problems are:

1. **Lack of explainability** - Most AI algorithms are like black-box algorithms, which give predictions without any visualization, making it difficult for the doctors to believe or check the prediction
2. **Lack of generalization** - Existing systems are trained using limited or homogenous datasets, thus lacking generalizability over diverse patient demographics and scanners.

3. **No integration into the clinical workflow** - Almost all proposed systems are standalone models, without any patient registration system, tracking the patient's history, and role-based management in hospitals.
4. **Class imbalance** - The minority classes of the Alzheimer's stages like ModerateDemented are significantly imbalanced in available data sets, making the algorithm unable to predict the minority classes.
5. **Computationally heavy** - Most of the systems need heavy computational power and GPU infrastructure, which is not feasible for hospitals with fewer resources.
6. **No longitudinal evaluation of disease** - None of the proposed systems have the capability of tracking the disease progression in multiple scans of patients over time.

1.3 Aim and Objectives of the Study

Purpose of the project is to develop an AI model that will be able to classify the stages of Alzheimer's disease in MRI scans of the brain. It will also provide some explanation of the output generated by the model.

Objectives:

1. Analyzing existing system
2. Developing a transfer learning based CNN to classify MRI scans into 4 different stages of Alzheimer's disease.
3. Implementing Grad-CAM to generate salient regions of the brain that play a role in prediction of the Alzheimer's disease stages.
4. Designing Hospital Management System which includes registration, upload of MRI scan, generating diagnosis report and history keeping.
5. Evaluate clinical relevance of the generated explanations for supporting medical decision making.

1.4 Significance of the Study

1. Early detection of Alzheimer's disease in VeryMild and Mild states provides an opportunity for medication and lifestyle modification
2. Incorporates transparency with the use of Grad-CAM heatmaps which help in knowing whether the algorithm is concentrating on the correct regions of the brain such as the hippocampus
3. Helps in better patient management by ensuring proper storage of scanned images, patient's history, and predictions using the Admin and Doctor modules, thus reducing the amount of paperwork in the busy Nigerian hospitals
4. Acts as a platform for further research where researchers can make use of it in diagnosing other diseases and other related problems
5. Helps in development of the health care sector in poor countries by running without the need of complicated hardware

1.5 Scope of the Study

The system operates solely on MRI detection. It categorizes images into the four stages of Alzheimer's disease based on the OASIS dataset. The hospital section consists of Admin and Doctor roles, which are responsible for registration, patient management, uploading images, making predictions, and reviewing patient history. Patients cannot log into the system.

Other forms of imaging such as PET scans, patient applications, or real-time application within the hospital setting are not included. The goal is to build a fully functional prototype that can be showcased during the final year project.

1.6 Limitation of the Study

There were certain limitations to this project. First, the CNN was trained based on the OASIS database, which is a great choice, but not always representative of all possible populations and scanning machines. Results may vary when using images from other hospitals.

The imbalance in classes (relatively few images of ModerateDemented category) is an issue that should be considered as it might influence the accuracy rate. This project is a prototype, hence further development and testing in a real-world environment are needed.

The visual explanation provided by Grad-CAM is helpful, but not always precise; sometimes heatmaps could appear overly wide. As the training needed GPU computing capacity (Google Colab), local hospitals without decent internet connection might require additional assistance.

As for the last limitation, currently there is no follow-up with patients involved in this project or any interface with hospital databasets yet.

1.7 Operational Definition of Terms

Alzheimer's Disease (AD): A progressive brain disorder that slowly destroys memory and thinking skills.

MRI: Magnetic Resonance Imaging: A safe scan that creates detailed pictures of the brain.

CNN (Convolutional Neural Network): A type of AI model especially good at analyzing images.

Transfer Learning: Using a model already trained on millions of images and fine-tuning it for brain scans.

EfficientNetB0: A modern, efficient CNN architecture used as the base model.

Grad-CAM: Gradient-weighted Class Activation Mapping a technique that creates heatmaps showing important image regions for a prediction.

XAI (Explainable AI): Methods that help humans understand why an AI made a particular decision.

Admin Module: The part of the system where a hospital administrator manages doctors and sees overall data.

Doctor Module: The part where doctors register patients, upload scans, and view results.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

Alzheimer's disease has steadily gained prominence among the most extensively researched neurological diseases in contemporary medicine due to the considerable impact that it has on memory, reasoning, behavioral patterns, and everyday life. The most frightening thing about this disease is that symptoms generally progress very slowly. In many instances, people are diagnosed with Alzheimer's only after considerable damage to the brain has been done. This has largely been one of the motivating factors behind scientists' ongoing search for improved diagnostic techniques (World Health Organization, 2023).

The use of neuroimaging technologies such as MRI and PET scans has been significant in the investigation of Alzheimer's disease over the years. MRI has become a common practice since it enables doctors to detect any structural changes to the brain non-invasively. Scientists have learned that areas like the hippocampus and medial temporal lobe tend to shrink in patients of Alzheimer's disease, even in stages before any serious symptomatology (de Leon et al., 1997; de Toledo-Morrell et al., 1997). However, interpreting an MRI scan is not easy for all cases.

In light of this, there has been an increasing application of Artificial Intelligence (AI) and Deep Learning to the field of medical imaging. Deep learning algorithms, especially CNNs, have demonstrated robust results in automatically identifying critical attributes from MRI images and categorizing various stages of Alzheimer's disease. Several researchers have successfully classified Alzheimer's disease using CNN-based models like ResNet, EfficientNet, DenseNet, and Vision Transformer on publicly available databases like ADNI and OASIS (He et al., 2016 ; Tan & Le, 2019 ; Dosovitskiy et al., 2021). These models are equipped to learn from thousands of MRI images which is not possible for conventional computing systems.

Even after the great success that deep learning models have achieved, there is another problem that exists. Most of the AI programs are just black boxes that do not give an explanation for their predictions. In a hospital environment, this can be a big problem since medical practitioners must trust the system before taking any decisions about treatment options. Without a clear explanation of how the predictions have been made, doctors will lose faith in the model, especially within the medical field (Samek et al., 2017).

Explainable AI (XAI) methods have emerged as a practical solution to this challenge. Techniques such as Grad-CAM, SHAP, and LIME each offer a different mechanism for rendering deep learning decisions more interpretable. Grad-CAM produces spatial heatmaps that indicate which areas of an MRI image were most influential in reaching a classification output, thereby allowing clinicians to verify whether the model is attending to neurologically meaningful regions (Selvaraju et al., 2017; Ribeiro et al., 2016; Lundberg & Lee, 2017).

The current chapter provides an overview of the theoretical background, technologies, and prior work relevant to the use of deep learning and explainable AI for Alzheimer's disease diagnosis. In particular, this chapter explores the pathogenesis of the disease, biomarkers on MRI images, and deep learning algorithms for processing of medical data, explainable artificial intelligence approaches, and recent findings in the field. At last, gaps in the existing research will be defined, which motivate this study.

2.2 Theoretical Background

2.2.1 Alzheimer's Disease: Definition, Epidemiology, and Pathophysiology

One of the major causes of dementia, Alzheimer's disease, accounts for 60-80% of all diagnosed cases around the globe. This is an irreversible and progressive brain disorder characterized by the gradual loss of cognitive functions such as memory, thinking, and behavior. Alzheimer's Disease usually occurs in adults aged over 65, although some early onset forms do occur.

The characteristic features of the disease include:

The appearance of extracellular amyloid beta (Aβ) plaques: The abnormal deposition of Aβ peptides outside the neurons.

Formation of intracellular neurofibrillary tangles (NFTs): Hyper phosphorylation and aggregation of tau protein inside the neurons.

Brain atrophy and neuron death due to the malfunctioning of synaptic functions.

This is followed by a certain spreading pattern of Alzheimer's Disease that includes first medial temporal lobe structures, including the entorhinal cortex and hippocampus, and further extends to the neocortex. As a result, the neurodegeneration process starts long before the onset of any symptoms.

The progression of the disease is staged using models such as the ATN system (Amyloid, Tau, and Neurodegeneration):

Preclinical/Asymptomatic: The presence of pathological characteristics without any cognitive impairment.

MCI: Amnestic MCI; many people progress to Alzheimer's Disease from this stage.

Dementia (AD): Difficulties with everyday activities (early, moderate, and severe).

2.2.2 Neuroimaging and MRI Biomarkers in Alzheimer's Disease

Neuroimaging becomes vital in detecting AD in non-invasive manner based on the visualization of morphological and functional brain alterations. While PET imaging can show amyloid/tau, MRI (particularly T1-weighted) is commonly employed for its availability, no radiation and possibility to quantify atrophy.

The main MRI-based biomarkers are as follows:

Hippocampal and medial temporal lobe atrophy, one of the best studied and earliest signs that correlate with cognitive impairment.

Enlargement of ventricles and thinning of the cortex in specific areas such as entorhinal cortex, temporal gyri and posterior cingulate.

Alterations in white and gray matter, quantifiable by means of volumetric measures or radiomics.

All above mentioned features correspond to stages of Braak staging of tau accumulation and allow discriminating CN, MCI and AD. MRI-based biomarkers allow establishing new criteria for diagnosing AD, thus making it possible to define the disease biologically even without dementia onset.

MRI remains to be the choice of many researchers due to its low cost and regular acquisition in routine practice although some problems such as inter-scan variance and automated image analyses remain relevant.

2.2.3 Deep Learning for Medical Image Analysis in AD Detection

Deep learning is an advanced machine learning technique used for automated feature extraction from complex inputs such as MRI images. While traditional machine learning approaches involve manually constructing the input features, DL algorithms leverage multi-layer neural networks for automatically extracting hierarchical representations of the inputs.

Popular neural network architectures for AD-MRI classification:

Convolutional Neural Networks (CNNs): Widely used for both 2D and 3D MRIs due to their inherent ability to model spatial hierarchies through the use of convolutional filters. Examples of popular CNN variants include ResNet, DenseNet, EfficientNet, and light models such as MobileNet. They are able to provide very accurate results (>85-99% accuracy) when used for binary (AD vs. CN) or multi-class (CN/MCI/AD) classifications of MRI datasets such as ADNI and OASIS.

Vision Transformers (ViTs): Novel architectures or hybrid techniques which view the images as a sequence of image patches and extract features through self-attention mechanisms. They are adept at modeling global interactions within MRI volumes and are usually superior or complementary to CNNs in current research (e.g., >98-99% accuracy in hybrid approaches).

2.2.4 Explainable AI (XAI) Techniques in AD MRI Classification

Though the deep learning models provide high levels of accuracy, they are black boxes in which it is tough for the physician to know about the process through which the decision is made, which can lead to a lack of trust among them. The problem is solved through the technique of Explainable AI or XAI, which explains in detail the whole process involved in prediction. XAI models can broadly be classified into two types depending upon the model: Gradient based (Model Specific) or Perturbation based (Model Agnostic).

Among the most common ones are Gradient-weighted Class Activation Mapping (Grad-CAM), Grad-CAM++, and HiRes-CAM. Grad-CAM helps in creating heatmaps by highlighting the crucial areas of the MRI image responsible for the decision made by the classifier. For instance, according to Junior et al. (2024) , in their multiclass AD classification model, which included LIME, the heatmap generated by Grad-CAM kept pointing to regions around the hippocampus and ventricles, thus improving predictability for medical professionals. Likewise, Alsubai et al. (2025) used a hybrid CNN-VGG16 model together with SHAP for several AD classification tasks from MRI images, achieving high accuracy scores along with consistent visual explanations matching AD-related atrophies.

Khan et al. (2025) created an IoMT-based framework with ResNet152 that made use of transfer learning, and employed different XAI techniques like Grad-CAM, SHAP, and LIME. Their model obtained a success rate of 97.77% on four classes of Alzheimer's dataset from Kaggle, and the results provided important insights on areas such as hippocampal and ventricular regions to gain more confidence about the model performance. Chakroborty et al. (2025) have conducted a thorough review of several methods of gradients (Grad-CAM, Grad-CAM++, HiRes-CAM, Backpropagation, and Guided Backpropagation) and model-agnostic methods on both pre-trained and custom CNN models used in AD prediction.

On the other hand, some of the prominent models in the field of perturbation include Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME), which are widely accepted due to their versatility, since they could be used with almost any machine learning and deep learning model. SHAP calculates the impact value assigned to each individual feature or voxel using game theory, and provides local as well as global explanation. In a study by Mahamud et al. (2025) , they employed an ensemble approach of LightGBM and Random Forest with SHAP and LIME, with an accuracy of 96.35%, and provided an identification of risk factors transparently.

One of the other techniques that can be employed is Layer-wise Relevance Propagation (LRP). This technique propagates the score of prediction backwards in order to assign a relevance score for each voxel or pixel. In their systematic review of XAI applied to Alzheimer's disease detection, Viswan et al. (2024) divided techniques into those which are post-hoc, ante-hoc, model-agnostic, or model-specific. These researchers pointed out that LIME and SHAP were among the most popular model-agnostic techniques in use, and showed that these techniques enhance clinical acceptance without significantly reducing accuracy. According to the review done by Khosroshahi et al. (2025) , the techniques of XAI can help bridge the AI-AD gap by generating relevant explanations.

Sheikh et al. (2025) concentrated on applying lightweight DL architectures in conjunction with XAI to develop a system suitable for regular clinical use. The research showed that the use of interpretable models makes it possible to implement the early diagnosis process effectively without increasing computational costs. Qin et al. (2025) . carried out consistency and stability benchmarking of Grad-CAM, SHAP, and LIME on Alzheimer's MRI data, noting the significance of testing the reliability and stability of the explanation results in various situations.

In general, current scientific papers illustrate that using both visual (such as Grad-CAM) and quantitative (such as SHAP) approaches yields better explanations. Nevertheless, there is little research conducted to create a model with intrinsic interpretability and check its stability, get an approval from clinicians, and validate the developed system on heterogeneous populations (Viswan et al., 2024 .0]; Khosroshahi et al., 2025 .; Khan et al., 2025).

2.2.5 Integration of XAI with Deep Learning for AD-MRI Analysis: Rationale and Challenges

Integration of DL models (such as CNNs/ViTs) into XAI results in highly accurate and explainable models for AD detection. Use of heatmaps and attribution methods proves that models pay their attention to meaningful areas, such as the hippocampus.

Difficulties:

1. Explanation consistency across different datasets.
2. Faithful evaluation of the explanation methods – how much explanations truly reflect models' internal logic.
3. Multimodal integration is required (including both MRIs and clinical features) and standardized XAI criteria.

Neuroimaging has become one of the main tools to explore and detect Alzheimer's disease, especially in its early stages. Based on research on this topic, MRI and PET scans are currently the most common imaging methods used in this case. In MRI, scientists are interested in changes in brain structure; in PET, in brain activity and metabolism.

As shown by researchers Norman Foster (2007) and Lisa Mosconi (2007), Alzheimer's disease causes both changes in brain structure (atrophy) and a decrease in glucose metabolism in certain regions of the brain. This change occurs over time gradually.

One particular fact that has been noted is that this process starts many years before the actual development of the disease, which makes it possible to predict the disease.

2.3 Review of Related Literature

An experiment carried out by Sheikh et al. (2025) involved lightweight deep learning models used in conjunction with Explainable AI in order to detect Alzheimer's disease at an early stage through regular MRI scans. The experiment made use of the ADNI dataset that included 2D MRI slices of 102 patients where the aim was to classify three types of cognitive stages including Cognitively Normal (CN), Early Mild Cognitive Impairment (EMCI), and Late Mild Cognitive Impairment (LMCI). The methodologies used in this study were lightweight neural networks in particular MobileNetV2 and EfficientNetV2B0 which were trained by employing transfer learning and data augmentation. In order to increase the interpretability aspect, Grad-CAM++ and Guided Grad-CAM++ were used in order to emphasize specific brain regions such as the hippocampus and the medial temporal lobe. The best model obtained had a mean cross-validation accuracy of 88%, increasing to 92% when data augmentation was used. However, limitations of this research include a small number of patients, only considering three stages of the disease rather than all four and lack of clinical application of the neuroimaging visualization software that was designed along with the models.

Pandey et al. (2024) conducted an analysis on how transfer learning is applied in deep convolutional neural networks in Alzheimer's disease detection to improve recommendation systems in healthcare. This research has been done using both ADNI and OASIS databases and some well-known models such as DenseNet-201, EfficientNet-B0, ResNet-50, ResNet-101, and ResNet-152. Preprocessing has been carried out using U-Net skull stripping which isolates brain tissue before classification. From the tested models, ResNet-101 performed the best and showed 98.21% accuracy on ADNI database and 97.45% on OASIS database in multi-class classification problems consisting of Alzheimer's Disease (AD), Cognitively Normal (CN), and Mild Cognitive Impairment (MCI). The paper proved that transfer learning using the modern convolutional neural network architectures has great prospects in clinical diagnostics. Nevertheless, this research suffered from the lack of Explainable AI implementation, validation in different scanners and different patients and from three-class classification instead of four-stage OASIS classification.

Mostafa et al. (2025) presented a novel hybrid deep ensemble model that distinguishes Alzheimer's Disease from Mild Cognitive Impairment using the ADNI data set. This methodology utilizes various transfer learning models — ResNet-50, NASNet, and MobileNet, employing a stacking technique for the meta-learner, thus enabling the ensemble to exploit the unique characteristics of the models used. Heatmaps and attribution maps were obtained using Grad-CAM in order to provide explanations for the decisions made by the model on gray matter and white matter slices. It was able to obtain a remarkable accuracy of 99.21% in distinguishing AD from MCI, and 91.02% when differentiating MCI from normal. Nevertheless, the study was limited by the inability of including four stages of differentiation instead of only two categories, as well as computational challenges associated with ensembles, along with the absence of detailed information regarding patients' specific data splitting.

Alsubai et al. (2025) proposed an architecture for a transfer learning deep learning system combined with Explainable AI to classify MRI images for various diseases such as brain tumors and Alzheimer's. This method uses the VGG16 network architecture with transfer learning, and the use of SHAP for the XAI approach, which provides visual explanations of the areas involved in making the model's decision. In conclusion, the study suggested that using transfer learning with XAI techniques increases the efficiency and interpretability of image classifiers in medicine. Yet, this research covered a wide scope involving different diseases; thus, it could not give enough information concerning Alzheimer's treatment compared to the focused studies. Besides, the VGG16 architecture is old and consumes more resources than newer approaches like EfficientNet; moreover, the paper lacked discussions regarding the implementation of their approach in actual hospitals.

Menon et al. (2024) have investigated an ensemble deep learning model that utilized Grad-CAM technique for visual interpretability in the case of Alzheimer's disease prediction. The authors used the dataset with MRI images divided into four stages of Alzheimer's disease and developed an ensemble model that integrated two independent networks: InceptionNet V3 and EfficientNet B7. Grad-CAM method was employed for visualizing and localizing the brain regions in the cases of four AD stages. An overall accuracy of 93% was reached by the ensemble model and was better compared to both architectures separately. However, the paper notes that ensemble architecture is much harder to implement because of its increased complexity. Besides, there is a possibility of overfitting without proper cross-validation; besides, 93% accuracy is somewhat lower compared to more recent researches in the field.

Chattopadhyay et al. (2024) performed a comparative study involving explainable AI models of Alzheimer's disease classification based on MRI imaging using an extensive data set that consisted of records from ADNI and NACC datasets, along with additional held-out data collected at NIMHANS in India. The method involved the use of DenseNet based CNN models supplemented by the application of both OSA (Occlusion Sensitivity Analysis) and Grad-CAM to create multi-scale explanations of decisions made by the model. The results showed that the combination of the two XAI approaches provided more complex explanations compared to each technique separately; however, in this study, more attention was paid to the comparative analysis of explainability methods rather than the maximization of classification accuracy and the explanations were not always highly detailed. Another deficiency of this study is the lack of clinical validation, which was demonstrated by a very limited participation of physicians in assessing the medical relevancy of the model explanations.

Aksouy and Daou (2025) created a web-based explainable diagnostic system for the diagnosis of AD using deep learning and XRAI explainability method. The database included a huge set of augmented MRI data, comprising 33,984 images from Kaggle divided into 4 levels of AD severity. The three deep learning models, namely MobileNetV3 Large, EfficientNet-B4, and ResNet-50, were trained and tested, while XRAI was used to generate visual region-level explanation for the predictions. The web-based application was created using Gradio library. The best result of 99.18% accuracy was reached using MobileNetV3 Large model with a weight of only 4.2 million parameters, which proves the idea of high performance and explainability capabilities in small models. However, the model was trained using only augmented Kaggle images, which may not completely cover all types of MRI scans, deployment of the application in limited resources hospital settings is problematic, and XRAI is less known in comparison with the commonly used Grad-CAM method.

In their research on IoMT-based Alzheimer's disease predictor using healthcare 5.0 paradigm, Khan et al. (2025) implemented a framework where transfer learning was combined with various XAI methods. The authors used an open-source Kaggle database consisting of 33,984 MRIs of four types of dementia, including Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. In their model, ResNet152 was used for feature extraction based on transfer learning, whereas Conditional Wasserstein GAN was utilized for dealing with imbalanced data and creation of synthetic images. As an explanation method, Grad-CAM, SHAP, and LIME were used, and, in all these approaches, attention was paid to clinically relevant brain regions like hippocampus and ventricles. An accuracy of 97.77% was achieved in the proposed model, and it was shown that a multi-method XAI approach improves the trust of physicians because of biological plausibility. Limitations of the study include usage of only one dataset from Kaggle that may lead to generalization problem, bias due to usage of generated images in GAN, and necessity for specialized hardware for IoMT.

Junior et al. (2024) explored the use of Explainable AI methods for multiclass Alzheimer's classification based on brain MRI scans from the ADNI database classified into Normal Cognition (NC), Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD) categories. They employed a ResNet-50 neural network enhanced with channel-wise attention to direct the model learning towards significant discriminant features. Both LIME and Grad-CAM visualization methods were utilized to identify critical brain regions contributing to decision making, with heat maps always indicating the hippocampus and ventricular regions, regions known to be associated with AD pathology. The use of attention mechanisms and combination of two Explainable AI techniques proved to enhance both accuracy and explainability of the model. Nevertheless, this work was limited only to three-class classification, used solely ADNI data and lacked clinical validation by doctors.

The Explainable Machine Learning Approach to Detecting Alzheimer’s Disease Using Ensemble Technique on a Structured Clinical Dataset was developed by Mahamud et al. (2025). In this case, their ensemble consisted of a combination of LightGBM and Random Forest classifiers, employing soft voting. Chi-square was applied to select the important features in the dataset. Afterward, SHAP and LIME were used to gain insight into how decisions made by the ensemble model were reached. The proposed system managed to achieve an accuracy rate of 96.35% on the test set, which is higher than the other competing models. However, the method is inherently non-image based; therefore, it is hard to make a comparison with MRI-based approaches. Also, its requirement for a structured clinical dataset means that it cannot be employed in early detection stages, when there is no clinical data available.

Viswan et al. (2024) have performed a systematic review to analyze the use of LIME and SHAP in Alzheimer's disease diagnosis in the scientific literature. Using the PRISMA and Kitchenham procedures for systematic review, the authors analyzed 23 articles that included different AD datasets and used LIME or SHAP in ML or DL models. Four categories of explanation methods post hoc, ante hoc, model agnostic, and model dependent are provided for comparative analysis. The findings supported the importance of using XAI techniques in making AI models more reliable and effective in medical practice, without compromising their prediction power. However, the study had some limitations since only two approaches to explainability—LIME and SHAP—were considered while ignoring other relevant techniques, including Grad-CAM and Layer-wise Relevance Propagation. Different sizes of data sets and differences in experimental designs prevented direct comparisons of the results. Furthermore, there was not enough involvement of healthcare specialists in the validation process of the clinical relevance of the XAI explanations.

In 2025, Khosroshahi et al. conducted a review on XAI approaches utilized in neuroimaging for diagnosing Alzheimer's disease. This review was based on the literature obtained from IEEE, PubMed, and Google Scholar through the utilization of several neuroimaging datasets. Among other methods, the research focused on the usage of SHAP, LIME, Grad-CAM, and Layer-wise Relevance Propagation (LRP) approaches in neuroimaging. It provided evidence of the capability of using XAI in order to discover biomarkers, predict disease progression and differentiate between stages of AD. It should be mentioned that XAI can become an efficient tool for connecting the outputs of AI algorithms and clinical practice. However, this approach has some limitations such as the lack of proper datasets, issues related to regulation, and the absence of metrics for assessing the quality of XAI. Besides, most of the models used in this study are characterized by post-hoc explainability that is not able to show the inner workings of AI algorithms.

The study by Qin et al. (2025) examined the consistency and stability of Grad-CAM, SHAP, and LIME in classification of both diffuse and focal brain MRI anomalies. The evaluation was performed using two different datasets: an augmented Alzheimer MRI v2 dataset, comprising of 6,400 slices in total from four dementia categories; and a Brain-Tumor MRI dataset consisting of 3,264 slices distributed among four tumor classes. All the XAI methods under analysis were applied using the same backbone 3D ResNet-50 with the same pre-processing steps involving skull stripping, MNI-152 space registration, and Z-score normalization. Stratifed patient-level sampling was done in order to avoid potential data leakage between train and test sets. Grad-CAM method was shown to perform better in terms of consistency and reliability of detecting both hippocampal atrophy and tumors. Deep-SHAP provided more anatomically rich explanations but is computationally expensive, and LIME had highly variable results depending on the explanation run. The limitation of this study consists of the restricted usage of one specific type of backbone architecture limiting the generalizability of results and augmented rather than clinical MRI images used in the Alzheimer's dataset.

Numerous works have investigated the application of AI technologies in the diagnosis of Alzheimer's disease. For example, CNNs like VGG16 and ResNet50 have achieved accuracy greater than 80%. This is due to the ability of these models to automatically learn relevant features of the MRI image without any human effort.

Nevertheless, despite their high performance, these models are commonly accused of being black boxes. To resolve this issue, the researchers have developed methods like Grad-CAM that show critical areas of the brain.

In another work, researchers have used LIME to elucidate the reasons behind the model's predictions. However, in some cases, the model indicated irrelevant areas.

After analyzing these studies, it becomes evident that even though the AI technology is effective, there is still room for improvement.

2.4 Summary of Related Literature

Table 2.1: Summary of Related Literature

AUTHOR	CONTRIBUTION	METHODOLOGY	LIMITATIONS
Sheikh et al. (2025)	Developed lightweight DL models with XAI for early AD detection from standard structural MRI scans, achieving 92% accuracy with visual explanation of hippocampal regions.	MobileNetV2 and EfficientNetV2B0 with transfer learning; Grad-CAM++ and Guided Grad-CAM++ for explainability; ADNI dataset (3-class: CN, EMCI, LMCI).	Very small dataset (102 patients); only three-class classification; no real-world clinical validation.
Pandey et al. (2024)	Compared five CNN architectures using transfer learning on ADNI and OASIS for AD and MCI detection; ResNet-101 achieved 98.21% on ADNI and 97.45% on OASIS.	DenseNet-201, EfficientNet-B0, ResNet-50/101/152 with transfer learning; skull stripping via U-Net; Gaussian filtering and contrast enhancement preprocessing.	No XAI integration; limited to three-class classification; no cross-scanner generalizability validation.
Mostafa et al. (2025)	Proposed a hybrid deep ensemble model with Grad-CAM for differentiating AD from MCI; achieved 99.21% accuracy for AD vs MCI classification on ADNI.	Ensemble of ResNet-50, NASNet, and MobileNet with stacked meta-learner; Grad-CAM heatmaps on gray and white matter MRI slices.	Binary classification only; resource-intensive ensemble; patient-based data split not clearly specified, risking data leakage.

Alsubai et al. (2025)	Built an explainable transfer learning framework for classifying brain tumour and Alzheimer's MRI images across three datasets, combining VGG16 with SHAP explanations.	VGG16 CNN with transfer learning; SHAP as the primary XAI method for visual attribution across multiple brain disease classification tasks.	Multi-disease focus reduces AD-specific detail; VGG16 is computationally heavier than newer architectures; limited discussion on real hospital deployment.
Menon et al. (2024)	Developed an ensemble deep learning system with Grad-CAM visualization for four-class AD staging from MRI, achieving 93% accuracy while improving prediction transparency.	Ensemble of InceptionNet V3 and EfficientNet B7; Grad-CAM for visual explanation of classification decisions across four AD stages.	High deployment complexity and inference time; risk of overfitting without careful validation; 93% accuracy trails more recent methods.
Chattopadhyay et al. (2024)	Compared XAI methods (Grad-CAM and OSA) applied to CNN-based AD classification on ADNI, NACC, and an Indian cohort (NIMHANS), achieving 91.3% cross-site accuracy.	DenseNet-based CNN with transfer learning; occlusion sensitivity analysis (OSA) and Grad-CAM for multi-scale explanation; leave-sites-out cross-validation.	Explanations can be coarse; focus on comparison rather than maximising accuracy; limited clinical validation with practising doctors.

Aksoy & Daou (2025)	Built a clinically deployable web-based diagnostic system with XAI for four-class AD detection from MRI, using MobileNetV3 which achieved 99.18% accuracy with real-time predictions.	MobileNetV3 Large, EfficientNet-B4, and ResNet-50; XRAI for region-based attribution; Gradio web interface; 33,984-image augmented Kaggle dataset.	Validated only on augmented Kaggle data; deployment challenges in low-resource settings; XRAI is less clinically established than Grad-CAM.
Khan et al. (2025)	Proposed an IoMT-enabled AD detection framework using ResNet152 and multi-method XAI, achieving 97.77% accuracy and highlighting hippocampal and ventricular regions.	ResNet152 with transfer learning; Conditional Wasserstein GAN for class imbalance; Grad-CAM, SHAP, and LIME for explainability on Kaggle four-class dataset.	Single dataset testing limits generalizability; GAN-generated synthetic images may introduce bias; IoMT hardware deployment not validated.
Junior et al. (2024)	Applied attention-enhanced ResNet-50 with LIME and Grad-CAM for multiclass AD classification on ADNI MRI data, consistently pointing to hippocampal and ventricular regions.	ResNet-50 with channel-wise attention mechanism; LIME and Grad-CAM for XAI; ADNI dataset with three-class classification (NC, MCI, AD).	Three-class only, not full four-stage AD spectrum; ADNI-only dataset limits cross-dataset generalizability; no clinical validation by practicing physicians.

2.5 Research Gap

Although deep learning models have achieved high accuracy in Alzheimer's disease diagnosis, several challenges remain, including the lack of transparency due to the black-box nature of these models, limited validation of XAI techniques by clinicians, poor generalization caused by reliance on single-modality or homogeneous datasets, and the high computational cost of many explainability methods. Additionally, existing studies often fail to provide patient-specific interpretable explanations, integrate multimodal data effectively, or support longitudinal analysis of disease progression. To address these limitations, this project combines deep learning and Explainable AI (XAI) within a web-based system to provide accurate, transparent, and clinically useful multi-class Alzheimer's disease classification from MRI data.

CHAPTER THREE

SYSTEM ANALYSIS AND DESIGN

3.1 Introduction

Now let us dwell on implementation of this system. In Chapter Three, we will start implementing our ideas by looking at what is happening in hospitals nowadays, what does not work properly and providing an insight into the way that the new innovative solution based on artificial intelligence technology will help in overcoming all these challenges.

It is necessary to explain that most of hospitals use doctors who analyze the patients' MRIs manually and document information in paper-based patient charts. Of course, this technique works perfectly, however, it has several weaknesses. The suggested system will utilize artificial intelligence technologies, provide explanation, and organize all processes in one single system. In addition, there are only three major objectives which serve as a background to our discussion in this particular chapter. They are aimed at solving the problem efficiently and developing innovative solutions.

In this chapter, analysis, requirements, design decisions, and integration will be discussed. Moreover, it will be possible to find all details which are necessary for understanding of this problem. For sure, I did my best to provide easy-to-understand information. Later, it will be supplemented with Use Case diagram, Activity diagram, and Class diagram.

3.2 System Analysis

3.2.1 Analysis of the Existing System

The classical procedure for diagnosing Alzheimer's disease in most hospital settings follows a structured multi-step pathway:

1. **Patient Presentation:** The patient presents with memory-related complaints and the physician conducts a clinical interview with the patient and a family member to establish the onset and progression of symptoms

2. **Cognitive Screening:** Standardised tools such as the Mini-Mental State Examination (MMSE) and Montreal Cognitive Assessment (MoCA) are administered to objectively measure memory, attention, language, and problem-solving ability
3. **Physical Examination:** A neurological examination is conducted to rule out other conditions that can produce dementia-like symptoms such as Parkinson's disease and vascular dementia
4. **Blood Tests:** Laboratory tests are ordered to exclude treatable causes of cognitive decline including thyroid dysfunction, vitamin B12 deficiency, and infections
5. **MRI Scan:** An MRI is prescribed and the radiologist examines the output looking for structural indicators such as hippocampal atrophy and ventricular enlargement, then produces a written report
6. **CSF Analysis:** Where available a lumbar puncture is performed to measure Alzheimer's biomarkers including amyloid beta and phosphorylated tau in cerebrospinal fluid
7. **Physician Consultation:** The doctor reviews all findings and consults with the patient and family to communicate the diagnosis and discuss management options
8. **Record Keeping:** Patient records including the MRI report, test scores, and clinical notes are documented, typically in paper form or through fragmented software systems with no unified interface

There are still hospitals where patient registration, MRI keeping, and reporting happen through different programs. The image is uploaded separately while all the reports and the predictions history of the patients are kept on paper.

3.2.2 Limitations of the Existing System

In the present situation, several issues arise. The first one is that the procedure takes too much time since there may be delays between receiving the results of the MRI and obtaining the definitive diagnoses. Secondly, the process is prone to errors due to the possible misinterpretation of the findings by the radiologist due to their minute nature. Thirdly, currently, there is no option of an additional interpretation from the technological side. Two professionals can disagree on the analysis of one image.

Another challenge is that when the use of new technologies is involved, there will be no explanations for why a patient gets particular results. At the moment, even though some hospitals apply AI, the answers provided do not give any reasons for the diagnosis received. It is challenging to trust such results entirely. Moreover, patients' data about the history of previous MRI scans, as well as predictions, are dispersed, making the process of monitoring more complicated. There are also resource-limited settings where the lack of specialists leads to delays in care.

3.2.3 Analysis of the Proposed System

The core objective of this project is to develop an accurate Convolutional Neural Network (CNN) model using Transfer Learning to classify MRI images into four stages of Alzheimer's disease: NonDemented, VeryMildDemented, MildDemented, and ModerateDemented. To enhance transparency and trust in the diagnostic process, the system integrates Explainable Artificial Intelligence (XAI) through Grad-CAM, enabling doctors to visualize the specific brain regions that influence the model's predictions. Additionally, a web-based hospital management system is developed with Admin and Doctor modules to facilitate patient registration, MRI upload, result visualization, and medical history management. Together, these components form the foundation of the system by combining accurate AI-powered diagnosis, interpretable decision-making, and a practical clinical workflow suitable for everyday hospital use.

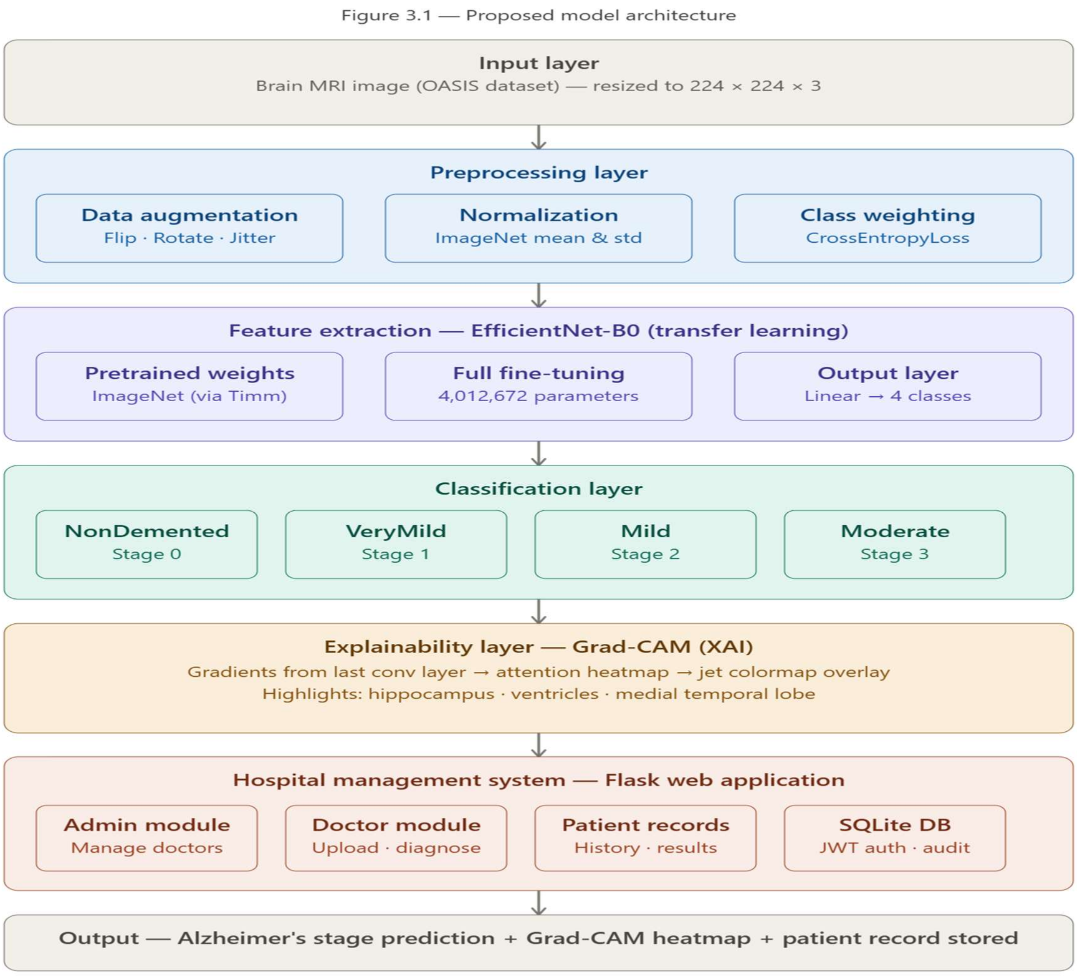


Figure 3.1 Proposed Model Architecture

3.2.4 System Requirements Specification (SRS)

A. **Functional Requirements:** It is important that the system be able to do certain actions. First of all, there needs to be an admin that logs into the system, creates unique codes for each doctor, works with doctors' accounts (activates/deactivates them), sees all patients and moves patients from one doctor to another.A Doctor should be able to register using the admin code, log in, add new patients (their name, age, gender, telephone number), choose the patient, send an MRI scan, and receive the predicted stage of the disease and its confidence, as well as the Grad-CAM map. It is also necessary to have access to the history of all scans and predictions of a particular patient.The AI module needs to take an

MRI scan, preprocess it, then run it through the efficient net B0 network, predict the stage of the disease with a confidence percentage and show a visual heatmap. All results need to be saved automatically.

B. **Non-Functional Requirements:** The system must be fast; prediction should come in a matter of seconds. In addition, it should operate normally on ordinary computer systems in the hospital and should not require expensive computer hardware for its operation. Security must be ensured; user data must be secure using log-in passwords and access control systems so that only doctors and administrators are allowed access.The graphical user interface must be user-friendly and should have minimum graphics for ease of use. The prediction model must be efficient in predicting and providing explanation for prediction results to physicians. In addition, it must be scalable to support multiple simultaneous users. Lastly, maintenance must be easy.

3.3 System Design Methodology

Two methodologies were adopted for this project. Object-Oriented Analysis and Design (OOAD) was used for system design, providing a structured blueprint through Use Case, Activity, and Class diagrams that defined all user roles, interactions, and data relationships before development. Transfer Learning was adopted for the machine learning component, where EfficientNet-B0 pretrained on ImageNet was fine-tuned on the OASIS MRI dataset with all 4,012,672 parameters fully retrained to adapt to Alzheimer's-specific MRI brain patterns.

3.3.1 Justification for Methodology

OOAD selected since the system and all user roles, data relationships and system interactions were well modeled in the Use Case, Activity and Class diagrams, before coding anything. With the limited MRI dataset, the accuracy at epoch 1 was 84.84%, which allowed for the reduction in training time. Transfer Learning was chosen because it increased the accuracy to only 1 epoch (84.84%) and improve the generalization from using the weights of the Inception_v3 pre-trained model from ImageNet. Because of its lightweight, low computation, medical image classification usage in literature and no need to spend on buying expensive GPU infrastructural facilities, the EfficientNet-B0 architecture was specifically selected over heavier ones such as ResNet and VGG. Partially freezing the layers wasn't done because the MRI images vary significantly from the normal images found in the ImageNet, making it necessary to be able to effectively learn the images of the brain in an Alzheimer's disease patient throughout the network.

3.4 Method of Data Collection

The data collecting process included study of related hospital systems, reviewing of papers in the last six years on the implementation of AI in Alzheimer's research, and interviewing people who know the processes of the hospital administration. My main data source was the OASIS MRI data set, obtained from Kaggle. The requirement analysis includes finding requirements based on the needs in relation to the doctors, to the admin, as well as the needs of the hospital in relation to patients.

3.4.1 Dataset Description and Preprocessing

The dataset used for the analysis is the OASIS Alzheimer's MRI dataset available from Kaggle. The dataset contains MRI images of brains arranged in four classes for Alzheimer's progression:

Table 3.1: Dataset Class Distribution

CLASS	**TRAINING SAMPLES**
NonDemented	18,071
VeryMildDemented	14543
MildDemented	10,908

ModerateDemented	833

The data had been divided between the train and test folders, both of which contained the four-class subfolders. In total, there were 44,355 images for the training set, while 46,184 were found in the test set. The major issue was that the ModerateDemented class had a very imbalanced number of images relative to the NonDemented class, 833 vs. 18,071 respectively.

3.4.2Preprocessing pipeline:

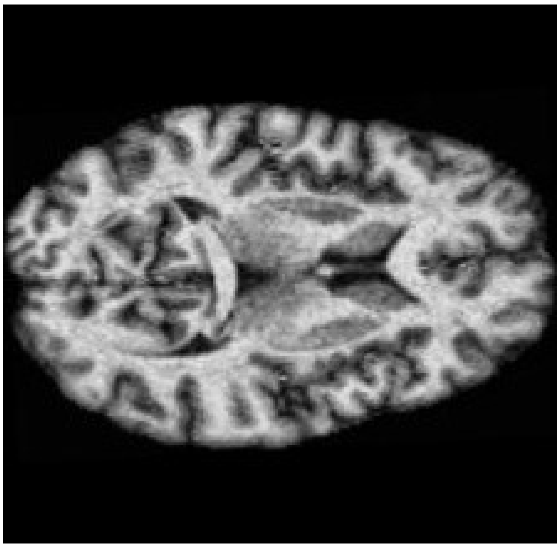
Before any image enters the model it goes through a fixed sequence of operations. For training images:

1. Resize to 224 × 224 pixels: the input size EfficientNet-B0 expects
2. Random horizontal flip: adds variation without changing clinical meaning
3. Random rotation up to 10 degrees: simulates slight head positioning differences in real scans
4. Color jitter with brightness and contrast variation of 0.2: helps the model generalize across different MRI scanner settings
5. Convert to tensor
6. Normalize using ImageNet mean [0.485, 0.456, 0.406] and standard deviation [0.229, 0.224, 0.225]

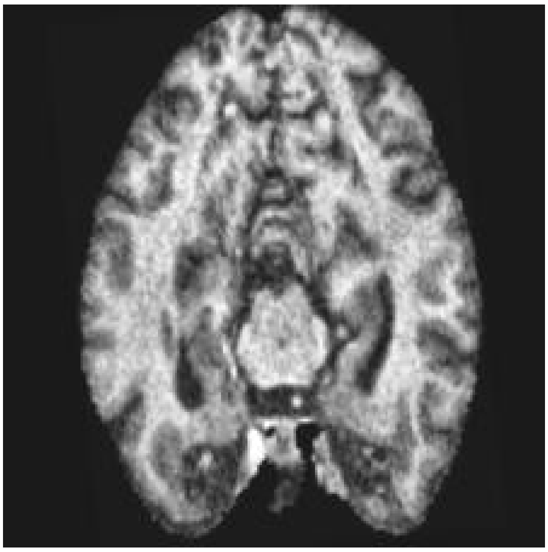
There was no other manipulation done to the validation and test sets other than resizing, tensoring, and normalization.

Further division of the training dataset yielded the training and validation sets at a ratio of 80/20. The resulting training and validation sets contained 44,355 images, distributed between 35,484 training images and 8,871 validation images, batched at 32 images per batch.

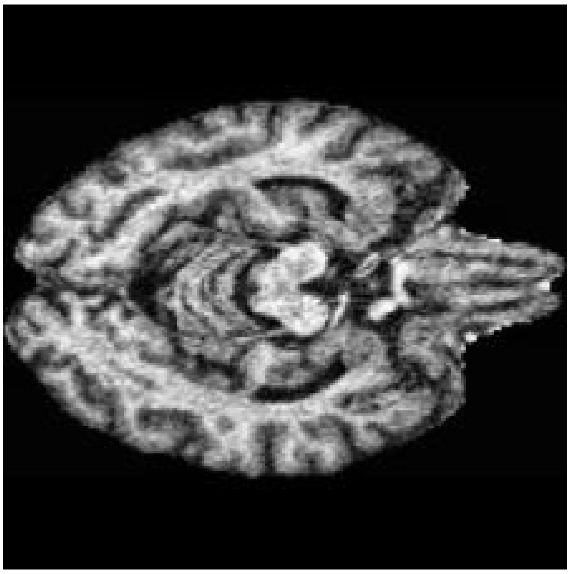
3.4.3 SAMPLE DATASET IMAGES



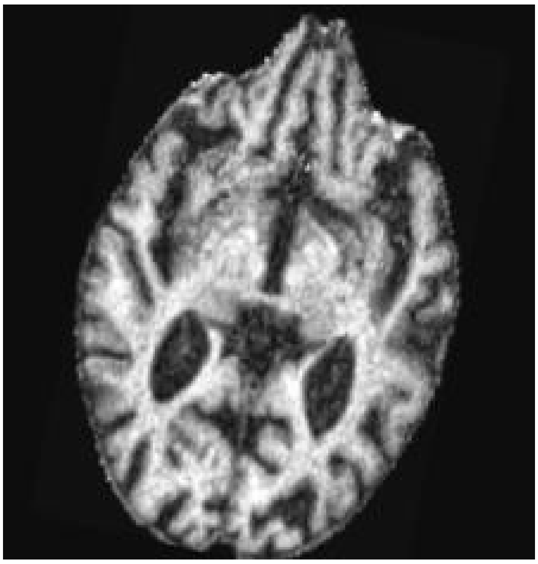
NonDemented



VeryMildDemented



MildDemented



ModerateDemented

Figure 3.2

Handling class imbalance:

Rather than manually resampling, class weights were computed automatically from the training data and passed into the CrossEntropyLoss function. The resulting weights were:

Table 3.2: Class Weight Computation

CLASS	WEIGHT
MildDemented	1.017
ModerateDemented	13.312
NonDemented	0.614
VeryMildDemented	0.763

The ModerateDemented weight of 13.312 effectively tells the model that misclassifying a ModerateDemented case is thirteen times more costly than misclassifying a NonDemented one. This directly compensates for the sample size disparity without needing synthetic data generation.

3.5 System Modeling

3.5.1 Use Case Diagram:

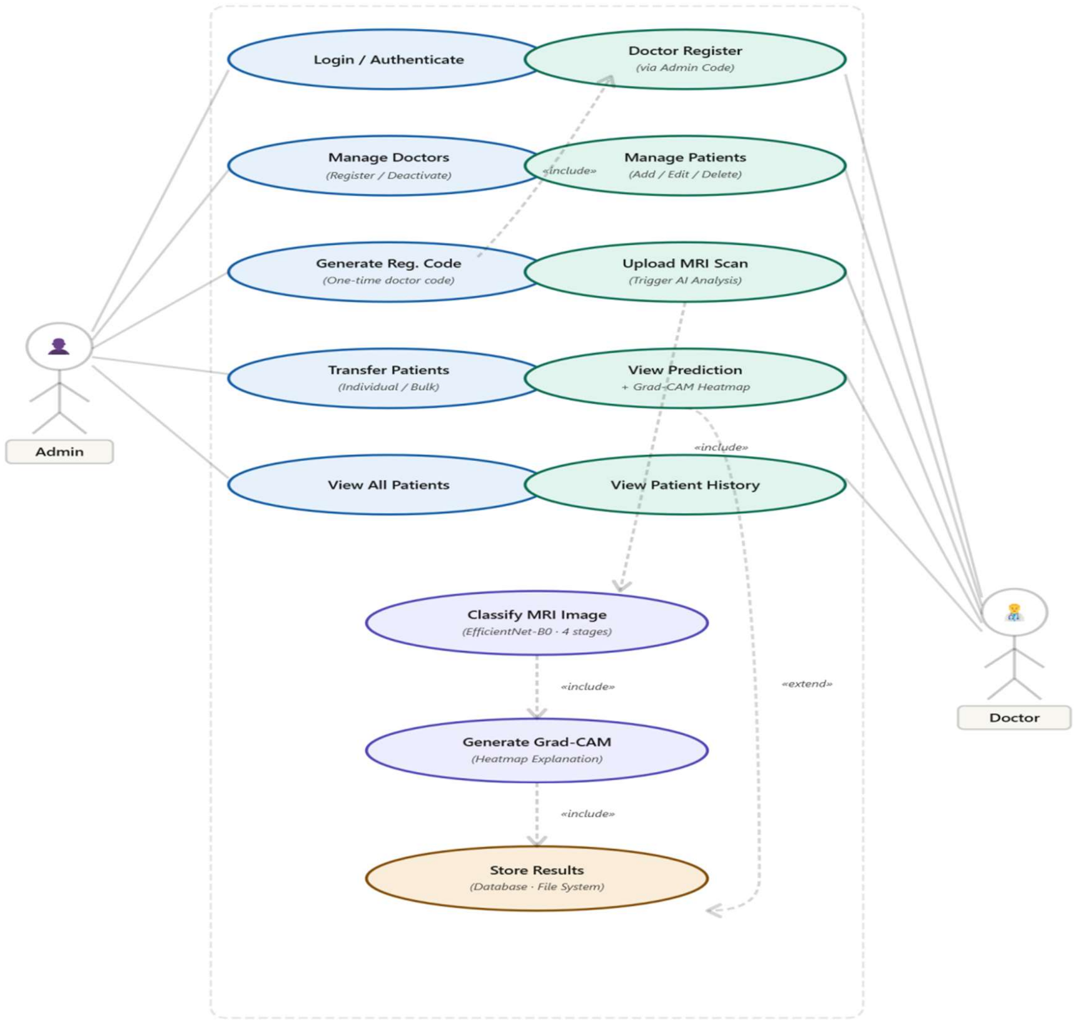


Figure 3.3: *Use Case Diagram*

The use case diagram shows the interactions between two actors, Admin, Doctor and the system. The Admin controls the system by adding doctor accounts, assigning and generate registration code, moving the patient to different doctors and seeing every record. The Doctor handles clinical operations by registering via an admin code, managing patients, uploading MRI scans, and viewing predictions with Grad-CAM heatmaps. The results become available in the database no later than 24 hours after the uploading of an MRI. Upon uploading of an MRI, 3 background operations are performed: MRI classification using EfficientNet-B0, Generation of heatmap using Grad-CAM and storage of the MRI results in the database through «include» relationships. The «extend» relationship between View Prediction and Store Results shows that data retrieved when a prediction is viewed is data that has been stored in the past.

3.5.2 Activity Diagram

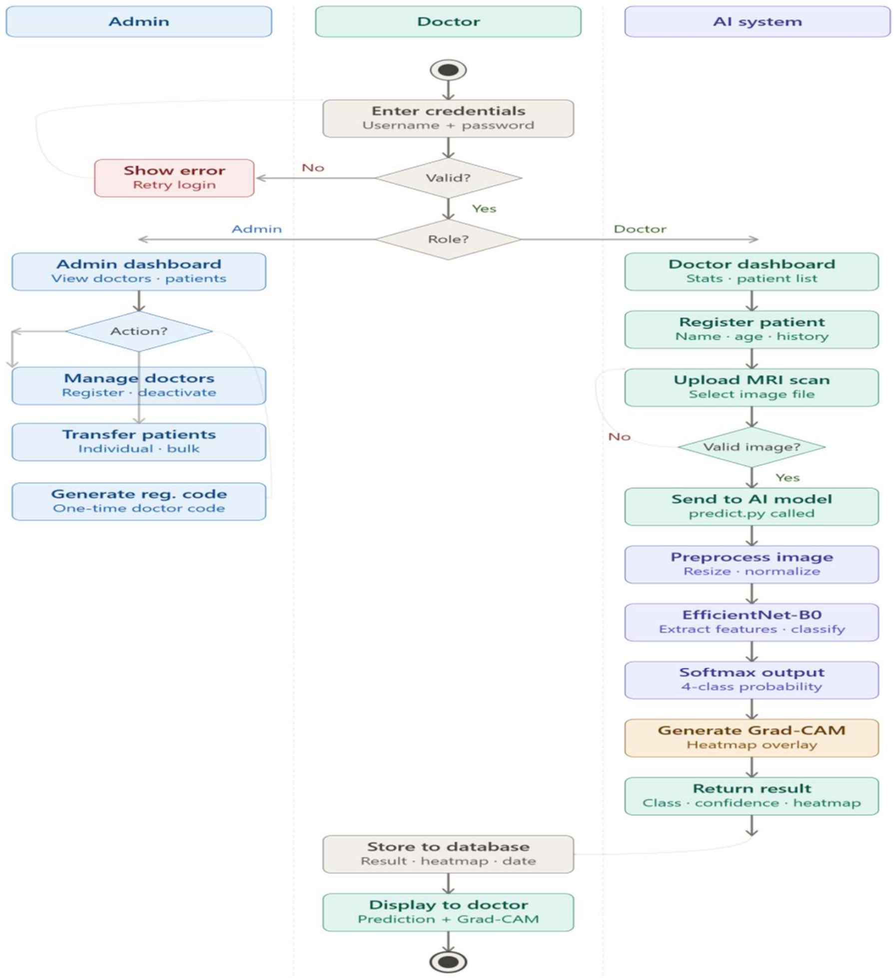


Figure 3.4: *Activity Diagram*

The activity diagram shows the full workflow of how the system operates and is broken down into three swim lanes: Admin, Doctor, and AI System. This is fired off by a shared login activity which verifies credentials, with bad credentials retried and good credentials doing an operation based on the user's role. The Admin lane is used to manage system operations, like doctor registration, patient transfers and one-time code generation. The Doctor lane includes key clinical workflows for patient registration, MRI upload and initiating AI analysis. The Doctor lane encompasses these essential clinical workflows, such as patient registration, uploading MRI scans, and initiating AI analysis. After getting a legal image, it moves to the AI System lane to preprocess the image, run it through EfficientNet-B0 to classify it, then run it through Softmax to get the probabilities of four classes, which comprises a heatmap, generated by a run-through of Grad-CAM. The returned result is put back into the shared zone, and remains in the database, where the doctor can view it.

3.5.3 Class Diagram

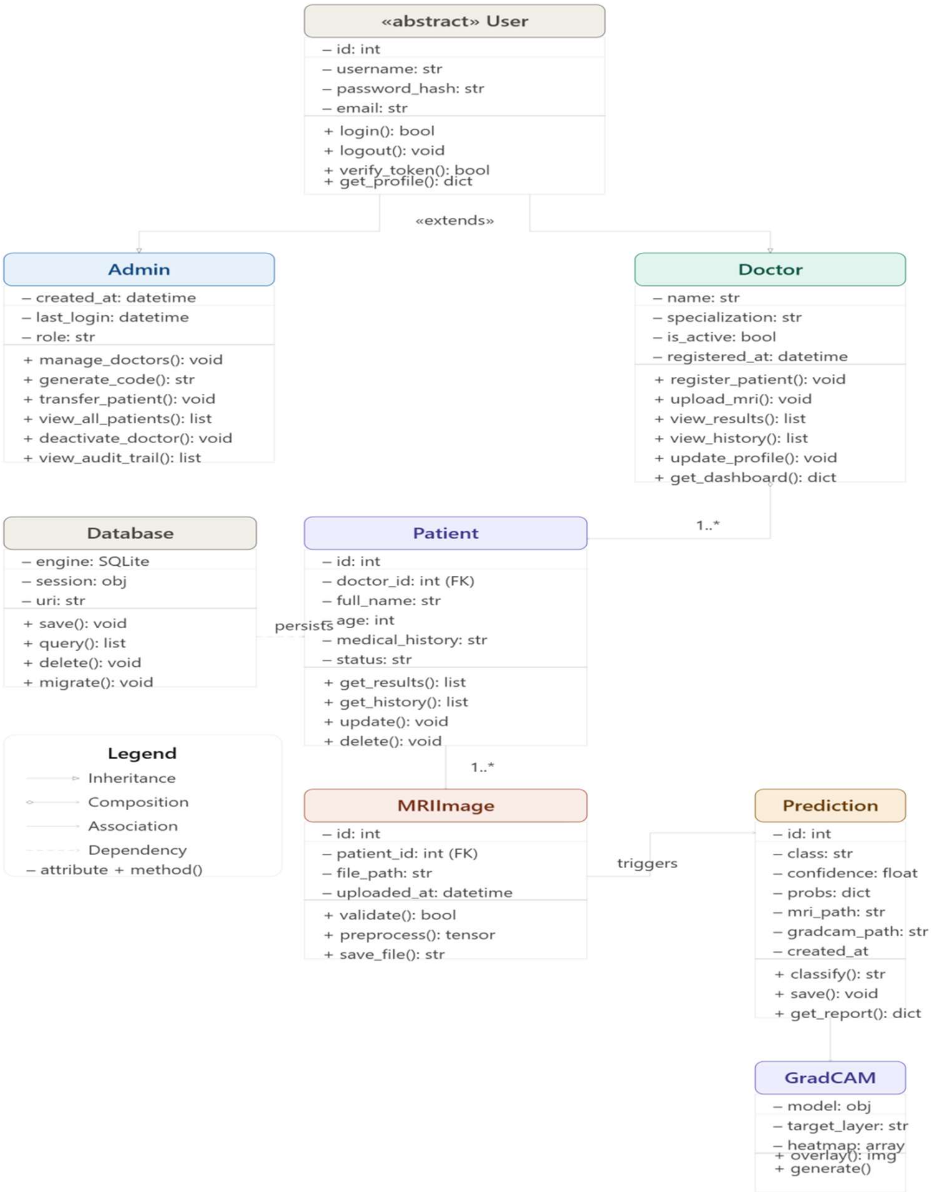


Figure 3.5: *Class Diagram*

The class diagram illustrates the object-oriented structure of the system. The abstract User class is used as a foundation for both Admin and Doctor; they have common attributes like ID, username, password hash, and email as well as common methods like log in, out, and validate token (indicated using inheritance relationship). Admin has user extension features for Doctor management, code generation and patient transfer; and Doctor has user extension features for patient registration, upload MRI, methods for viewing results. As a result, the Doctor class has a composition relationship with Patient - a patient cannot exist without being assigned to a doctor. Each patient can have many scans so the relationship of Patient to MRI Image is one to many. The Prediction class is triggered by MRI Image and carries out the EfficientNet-B0 classification and outputs the result and path of Grad-CAM heatmap. Prediction calls the Grad CAM class to perform and generate the heatmap overlaid on the original MRI. The Database class has also been separated out, and tightly coupled with the Patient class, meaning that all the persistence operations are executed through the Database layer, which stores all patient information.

3.6 System Design

3.6.1 Database Design

We use a relational database (SQLite for the prototype, easy to upgrade to PostgreSQL). Main tables include:

- Users (only for Admin and Doctors)
- Patients (linked to one doctor)
- Final Predictions which is linked to each individual patient.

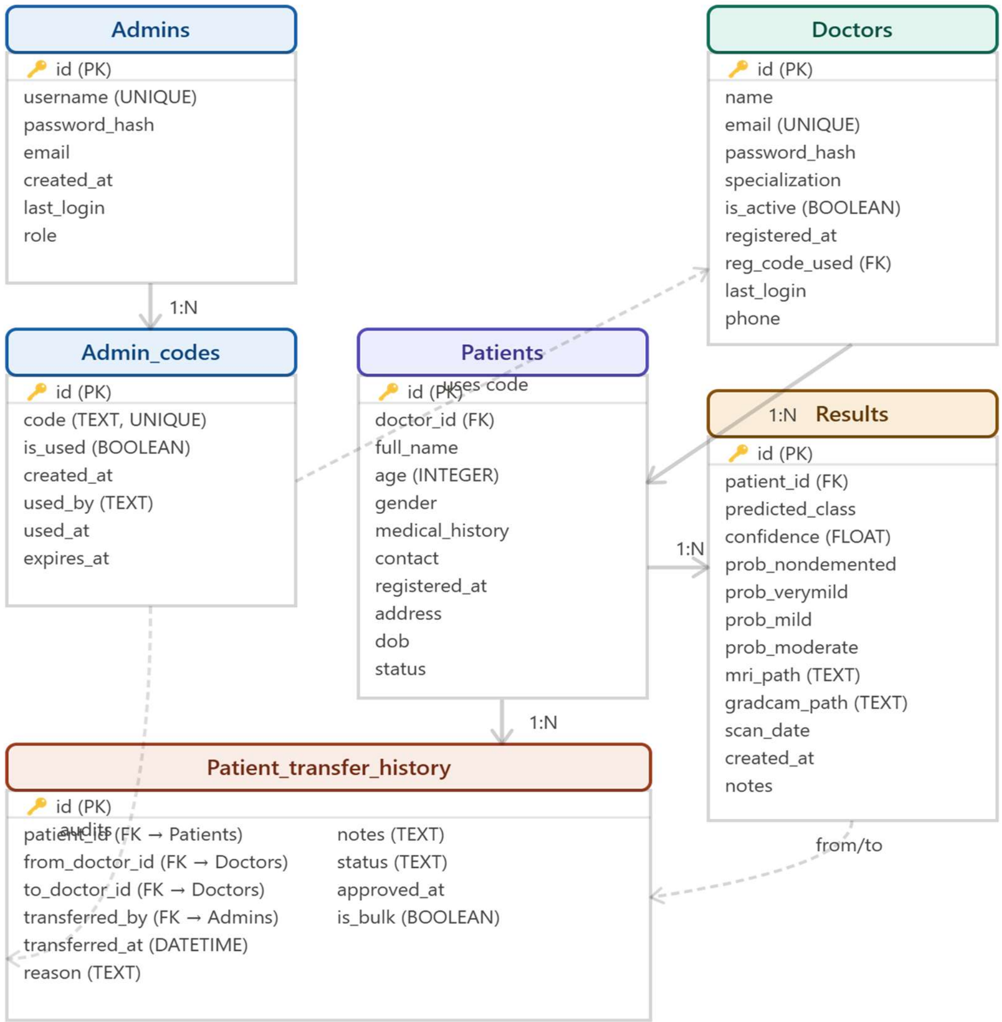


Figure 3.6: *Database Design*

The following database diagram depicts the 6 database tables making up the data layer of the hospital management system. Admins – table that contains session and admin credentials. The Doctors table contains doctor profiles and includes a foreign key to Admin_codes to maintain a relationship with the registration code that's used during sign up for the doctor. One-time registration codes are generated by a user (usually Admin) and stored in the Admin_codes table, where they can be checked for use for registration. Table Patients contains data on the patients, with a Foreign Key that connects each patient to the doctor they belong to. The Results table records all Alzheimer's predictions on a patient, with the predicted Alzheimer's stage, the confidence score, every probability for one of the four classes, and file names of the original MRI image as well as the heat map generated by Grad-CAM. The Patient_transfer_history table spans the patient's moving of information between doctors, capturing who authorized the transfer and the time the move was authorized, to offer a full audit path as to how the patient was transferred through the hospital.

3.6.4 System Architecture Design

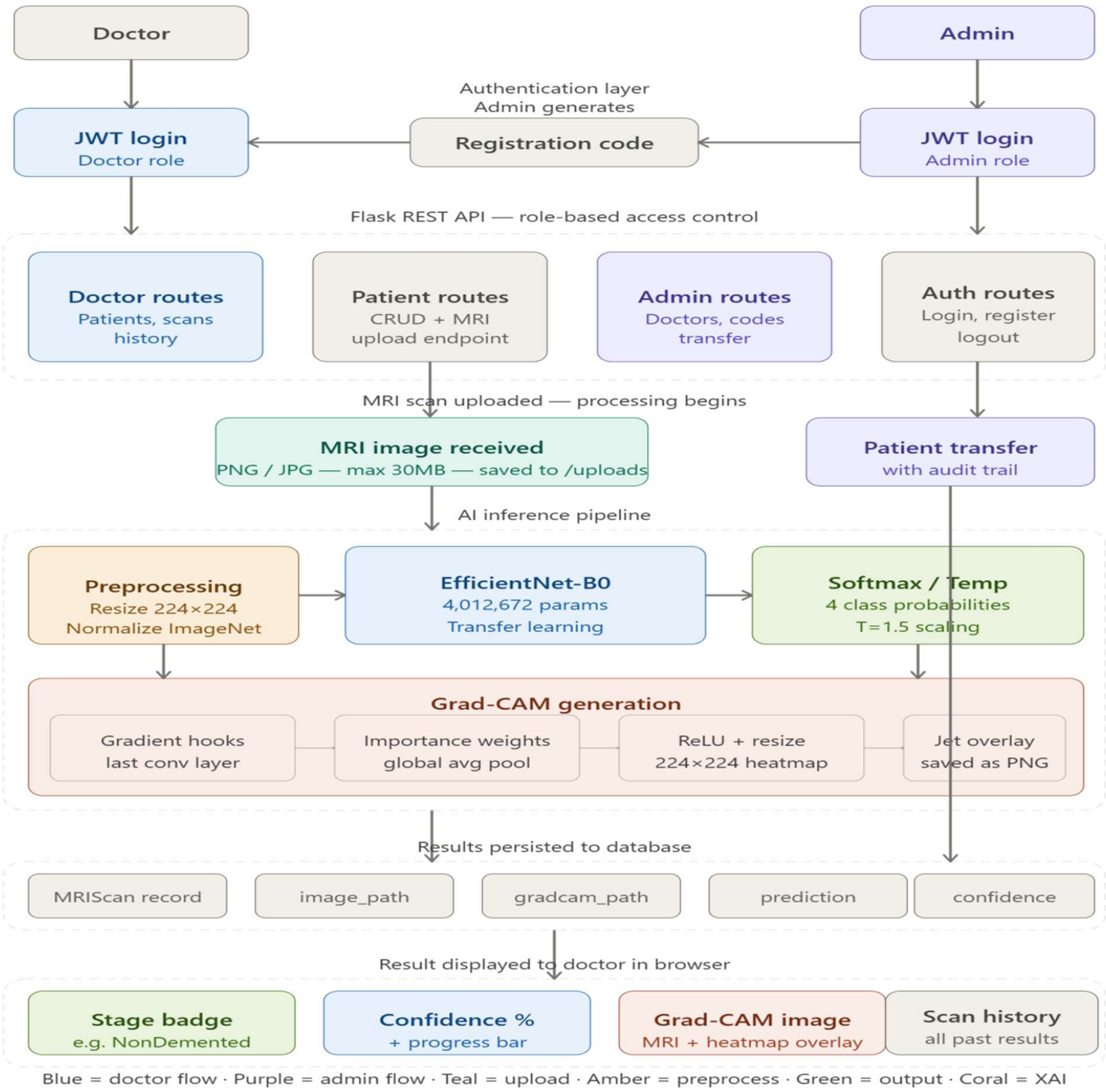


Figure 3.7: *System Architecture Design*

The diagram shows the end-to-end architecture of the AI-based Alzheimer's classification system and how data flows through the system. The Doctor logs in via JWT and the Admin generates a one-time registration code that the Doctor logs in with to register. Flask REST API role based access control works by separating their methods/routes once they log in, so The Doctor can get access to patient history, scan history and so on routes and Admin can get access to doctor management, patient transfer, and so on routes. On upload of an MRI scan by the Doctor an AI inference pipeline is triggered, the image is resized to 224x224 and normalized, then sent through EfficientNet-B0 (4,012,672 parameters with transfer learning) and the end result is four class probabilities using Softmax. Importance weights, obtained by computing a 224×224 heatmap, computing ReLU activations and applying the jet colormap to them using gradient hooks on the final convolutional layer, are then saved as a PNG, to be extracted and displayed later. The prediction results, such as the MRI path, Grad-CAM path, predicted class and confidence percentage with progress bar are stored in the database and shown in the browser as a stage badge, confidence percentage with progress bar, Grad-CAM heatmap overlaid on the MRI and the full scan history.

3.7 Model Architecture and Training

3.7.1 Model Architecture

EfficientNet-B0 architecture was used as the backbone for the project. It was selected based on three criteria: computational efficiency, enabling its usage even on CPU-based machines; successful application in recent papers for the task of medical image classification; and scalability in case the experiment would be scaled from CPU to GPU-based machine.

It was imported using ImageNet pretrained weights via the timm library. A linear classification layer outputting four nodes corresponding to four Alzheimer's stages was inserted instead of the last layer by default. Total number of parameters in the architecture equals 4,012,672, and all of them were tuned during the training process. Instead of freezing the backbone at first and training just the head, a joint tuning of the whole architecture took place, due to the significant domain shift.

This is how the forward process is done. The tensor containing one MRI scan of size (1, 3, 224, 224) is fed into the model through the compound-scale convolutional layers of EfficientNet. The output is four raw logit scores. The Softmax function normalizes the logit scores to obtain the probabilities which will always be one. The maximum probability corresponds to the prediction, and the value of the probability is returned.

The raw output of the model is a vector of four logit values. These are converted to class probabilities using the Softmax function:

$$P(y = i|x) = \frac{e^{z_i}}{\sum_{j=1}^{4} e^{z_j}}$$

Where:

- **z_i** = the raw logit output for class *i*
- **e** = Euler's number (base of the natural logarithm)
- The denominator sums over all four classes to ensure probabilities sum to 1.0

The predicted class is the one with the highest probability:

$$\hat{y} = \arg \underbrace{max}_{i} \; P(y = i|x)$$

And the confidence score returned to the doctor is simply:

$$Confidence = \underbrace{max}_{i} P(y = i|x) \times 100\%$$

3.7.2 Training Configuration

Table 3.4: Training Configuration Parameters

PARAMETER	**VALUE**
Optimizer	Adam
Initial Learning Rate	0.0001
Loss Function	CrossEntropyLoss with class weights
Batch Size	32
Epochs	25
Learning Rate Scheduler	SteLR: reduced by factor 0.1 every 7epochs

The training process employed checkpoints, which allowed saving of the model parameters at the end of each epoch. This enabled pausing and resumption of training while preserving the progress made thus far, which was crucial given the long training period on the CPU. The checkpoint with the best performance in terms of validation accuracy was saved to become the final model.

The loss function used during training is Categorical Cross-Entropy Loss with class weights to handle the dataset imbalance:

$$L = -\sum_{i=1}^{4} w_i \cdot y_i \cdot log(p_i)$$

Where:

- **w_i** = the computed weight for class *i*
- **y_i** = the true label (1 if the sample belongs to class *i*, 0 otherwise)
- **p^i\hat{p}_i p^i** = the predicted probability for class *i* from Softmax

The class weights themselves were computed as:

$$w_i = \frac{N}{K \times n_i}$$

Where:

- **N** = total number of training samples (44,355)
- **K** = number of classes (4)
- **n_i** = number of samples in class *i*

Table 3.5: Class Sample Sizes for Weight Computation

CLASS	n_i	Weight w_i
MildDemented	10,908	1.017
ModerateDemented	833	13.312

NonDemented	18,071	0.614
VeryMildDemented	14,543	0.763

The Adam optimizer updated model weights using the following update rule:

$$\theta_{t+1} = \theta_t - \frac{n}{\sqrt{\hat{v}_t + \epsilon}} \cdot \hat{m}_t$$

Where:

- **θ** = model parameters (weights)
- **η** = learning rate (0.0001)
- $\hat{m}_t$= bias-corrected first moment estimate (mean of gradients)
- $\hat{\boldsymbol{v}}_{\boldsymbol{t}}$ = bias-corrected second moment estimate (variance of gradients)
- **ε** = small constant to prevent division by zero (typically 1e-8)

3.7.3 Training Results

The model showed strong performance from the very first epoch:

Table 3.6: Training Results per Epoch

EPOCH	TRAIN LOSS	TRAIN ACCURACY	VALUE LOSS	VALUE ACCURACY
1	0.6590	74.11%	0.3274	84.84%

5	0.0430	98.19%	0.1105	95.52%
10	0.0024	99.96%	0.0451	98.12%
25	0.0007	99.99%	0.0412	Best checkpoint saved

This is because the gap between 74.11% training accuracy and 84.84% validation accuracy in just one epoch shows that the transfer learning using ImageNet pre-trained model was successful in MRI data. It is normal for the training accuracy to be lower than validation accuracy in the early epochs since the training images are augmented making them more difficult.

3.8 Testing and Evaluation

3.8.1 Evaluation Metrics

Four metrics were used to evaluate the model's classification performance:

1. **Accuracy**: the percentage of total predictions that were correct across all four classes. This gives the overall picture but can be misleading when classes are imbalanced, which is why it is not the only metric used.

   Accuracy measures the proportion of correctly classified images out of the total number of images evaluated across all four Alzheimer's stages.

   Accuracy = $\frac{TP+T}{TP+TN+FP+FN}$

For the multi-class case with four stages, this extends to:

Accuracy = $\frac{\sum_{1=1}^{4} Correctly\ Classified\ Samplesof\ Class\ i}{Total\ Number\ of\ Samples}$

Where:

- **TP** = True Positives (correctly predicted as a class)
- **TN** = True Negatives (correctly predicted as not a class)
- **FP** = False Positives (incorrectly predicted as a class)
- **FN** = False Negatives (incorrectly predicted as not belonging to a class)

2. **Precision:** of all the images the model predicted as a particular class, what percentage actually belonged to that class. High precision means fewer false positives.

   The precision for each class is calculated as:

$$Precision_i = \frac{TP_i}{TP_i + FP_i}$$

   The macro average precision across all four classes is:

$$Precision_{macro} = \frac{1}{4}\sum_{i=1}^{4} Precision_i$$

Where:

- **TP_i** = True Positives for class *i*
- **FP_i** = False Positives for class *i* (images from other classes incorrectly predicted as class *i*)

3. **Recall (Sensitivity)** — of all images that actually belong to a class, what percentage did the model correctly identify. This matters especially for ModerateDemented — missing a severe case is a serious clinical error.

Recall (also called Sensitivity) for each class *i* is:

$$Recall_i = \frac{TP_i}{TP_i + FN_i}$$

The macro-averaged recall across all four classes is:

$$Recall_{macro} = \frac{1}{4}\sum_{i=1}^{4} Recall_i$$

Where:

- **FN_i** = False Negatives for class *i* (images that belong to class *i* but were predicted as a different class)

4. **F1-Score** — the harmonic mean of precision and recall. This is the most informative single metric for imbalanced datasets because it penalizes models that sacrifice recall for precision or vice versa.

The F1-Score is the harmonic mean of Precision and Recall for each class *i*:

$$F1_i = \frac{Precision_i \times Recall_i}{Precision_i + Recall_i}$$

The macro-averaged F1-Score across all four Alzheimer's classes is:

$$F1_{macro} = \frac{1}{4}\sum_{i=1}^{4} F1_i$$

The harmonic mean is used rather than the arithmetic mean because it penalizes cases where either precision or recall is very low, which is important in medical classification where both missing a case and over-predicting are costly.

For imbalanced datasets such as the OASIS test set used in this study, the weighted average F1-score is the most appropriate summary metric as it accounts for the unequal distribution of samples across the four Alzheimer's disease stages.

5. **Confusion Matrix:** shows exactly where the model gets confused between classes. A well-trained model should show strong diagonal values and minimal off-diagonal

confusion, particularly between adjacent stages like MildDemented and ModerateDemented.

3.9 Results Presentation and Analysis

3.9.1 Output Samples

Training Progress Plot

The training progress plot is a graph that shows the training behaviour of the model at 25 epochs with four curves representing the validation loss, training loss, validation accuracy and training accuracy. The plot shows that both the loss and the accuracy declined rapidly during the initial 8 epochs, improved learning of the pretrained weights on the ImageNet data was noticed from the start, and it was adequate for 84.84% accuracy on valid loss data already within the first epoch. Starting from epoch 8, the two curves were seen to be very flat and became parallel towards the end, indicating no overfitting of the model. The accuracy of the training continually remained above 99.94% from 10th epoch onward, with the best accuracy obtained at the 19th epoch which was stored as a checkpoint for the checkpoints folder, attaining a 98.66% accuracy. The training objective was greatly decreased from 0.6590(epoch 1) to 0.0007 (epoch 25), thus achieving a high level of convergence in the model.

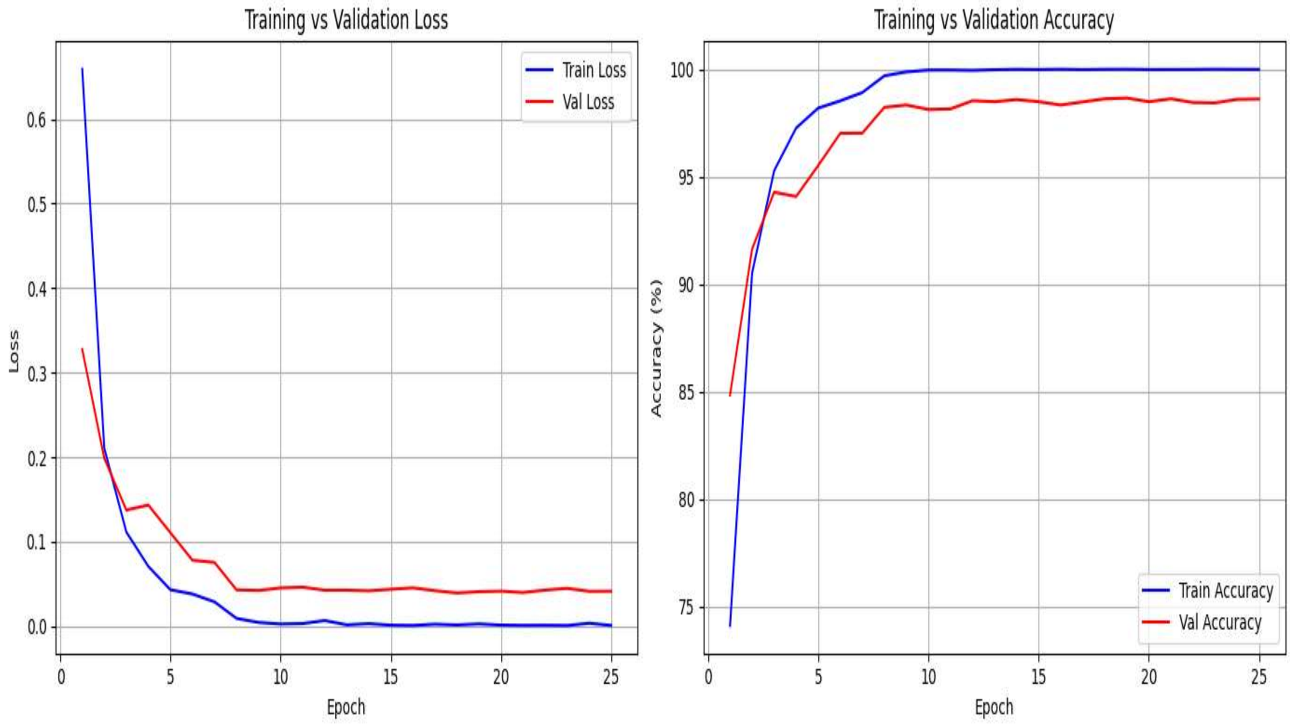


Figure 3.8: *Training Progress Plot*

Confusion Matrix

The confusion matrix is the summarization of the accuracy of the model at each of the 4 stages of Alzheimer's disease on a test set of 46,184 images. The cells that are active diagonally indicate the ones which were predicted correctly and off diagonal cells are those in which prediction was incorrect. As expected, NonDemented had the highest precision of .9550 and .8100 recall with a dominance of 88.5% on the test set. Due to the moderate value of 0.5463 in terms of recall, the model labeled results to about one-half of actual Very Mild patients. On the whole, MildDemented has lower precision (0.0671) and recall (0.2610) and most cases were misclassified due to similar brain structural features at the beginning of the disease. ModerateDemented, therefore, had the poorest performance due to the fact that it lacked 168 out of 46,184 test samples in this class, resulting in no accuracy or recall. The overall weighted F1-score of 0.8126 indicates that the NonDemented are the classes performing well, but the macro F1-score of 0.3374 illustrates the poor performance on minority classes.

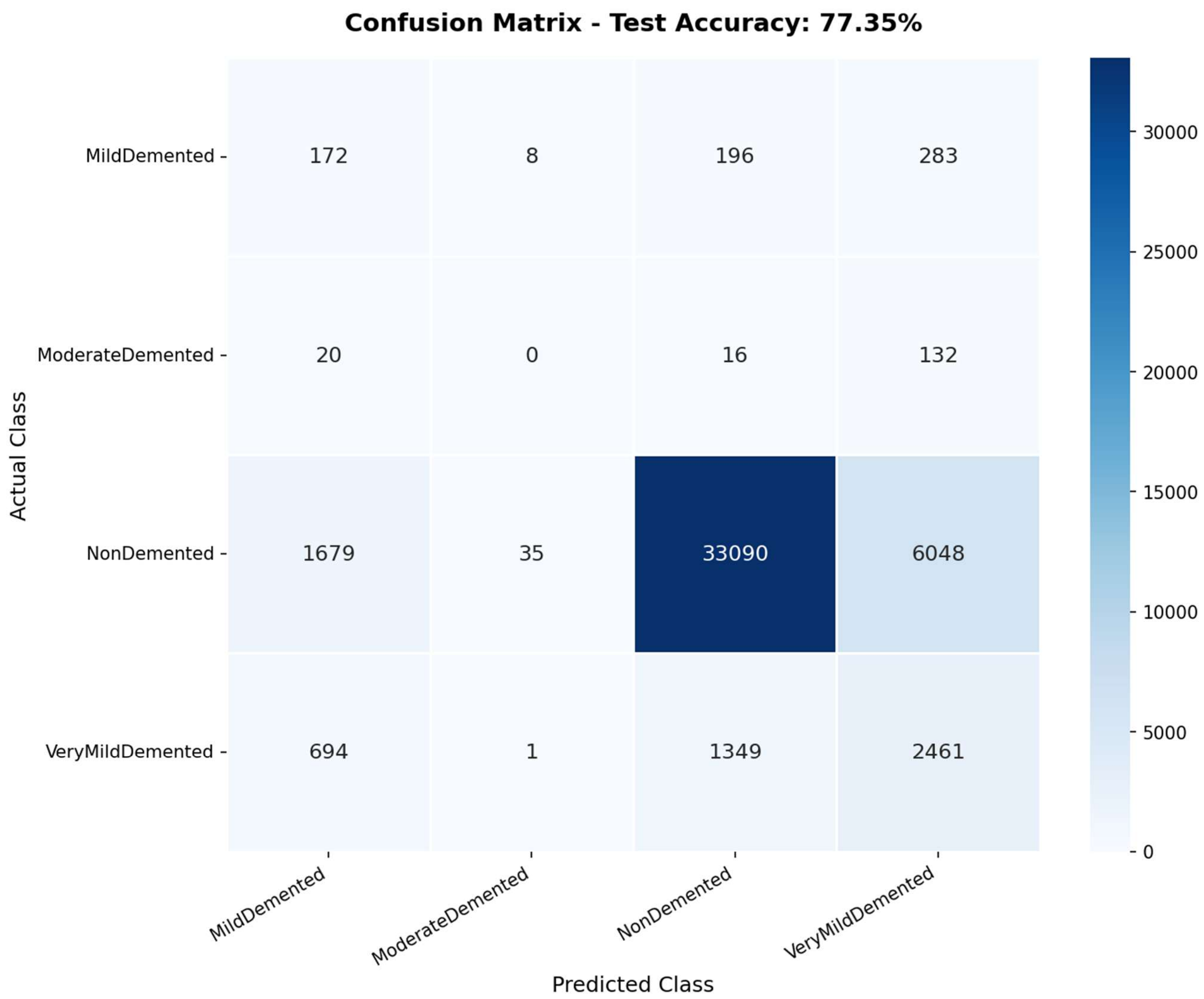


Figure 3.9: *Confusion Matrix*

ROC CURVE:

The ROC (Receiver Operating Characteristic) curve is used to compare how well the model categorizes each of the Alzheimer's classes from the other classes using One-vs-Rest class probability, with True Positive Rate vs False Positive Rate trade off assessment at increasing levels of threshold. The closer the curve to the top left corner, the better the discriminative powers, and a diagonal line will mean that a random guess at the class labels would be as good. The NonDemented class had the largest proportion, resulting in the best AUC performance when compared to other classes, due to its large proportion. VeryMildDemented and MildDemented had moderate AUC values, suggesting that the model was able to differentiate between the two classes, but with some overlap. The class of "ModerateDemented" had the lowest AUC and showed the least recall in the confusion matrix due to the model's low predictive ability of this class. Generally, the ROC observed that the accuracy of the majority classes is high and the accuracy of the minority classes is low, so it is necessary to find other appropriate class balancing strategies in further research.

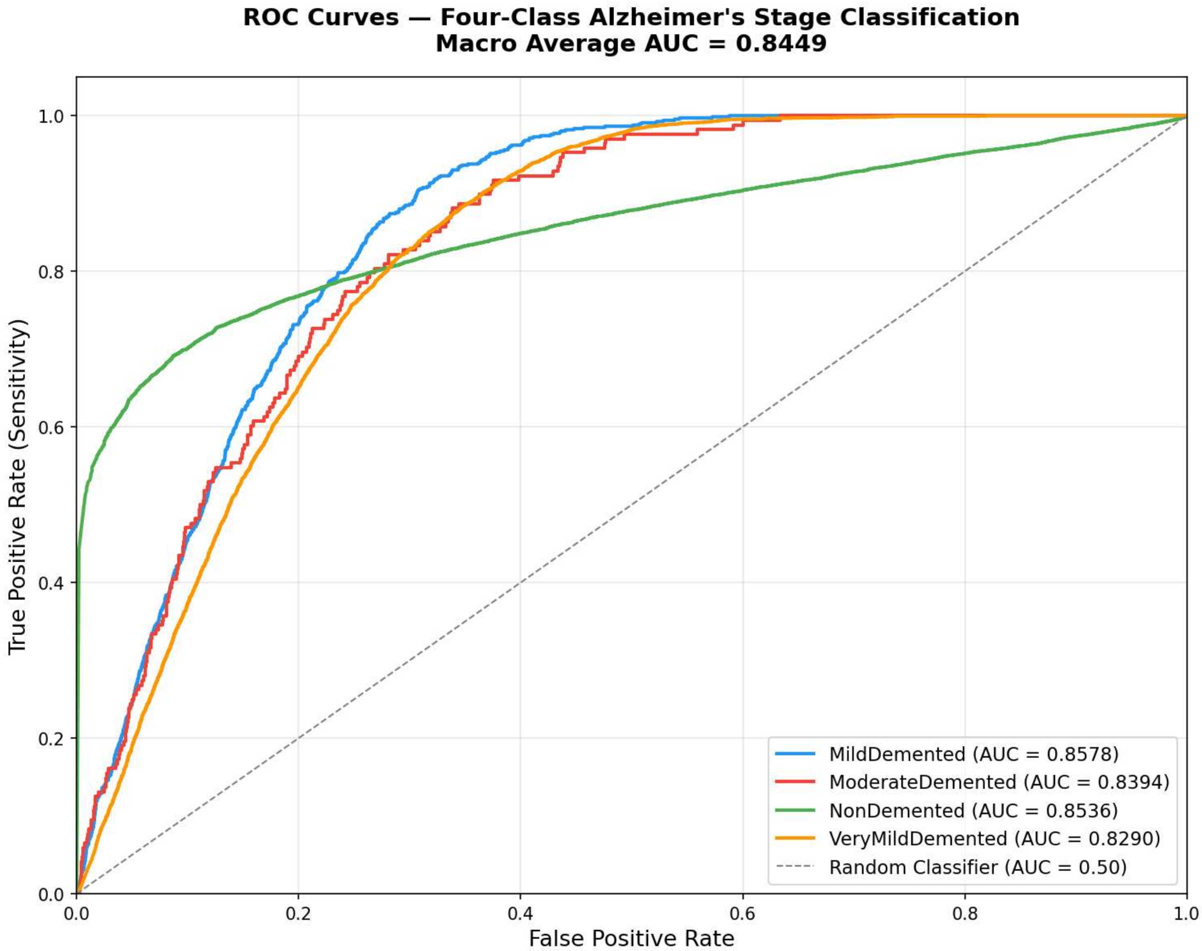


Figure 3.10: *ROC Curve for All Alzheimer's Disease Classes*

3.9.2 Discussion of Results and System Performance

Training Performance

The model exhibited consistent performance with respect to learning in all of the training epochs, that is, for all 25 epochs. The learning began from training accuracy of 74.11% and validation accuracy of 84.84% in the first epoch, and the model performed excellently during the initial eight epochs because of the feature transfer from the pretrained ImageNet to MRI datasets. In the 8th epoch, validation accuracy became 98.22%, while the best accuracy in validation phase was 98.66%, which was achieved during the 19th epoch. Overall training time was 7,153.28 minutes, about 119 hours. The full training progression is summarized below:

Table 3.7: Full Training Progression per Epoch

EPOCH	TRAIN LOSS	TRAIN ACCURACY	VAL LOSS	VAL ACCURACY
1	0.6590	74.11%	0.3274	84.84%
2	0.2104	90.51%	0.1985	91.63%
3	0.1113	95.27%	0.1373	94.27%
4	0.0706	97.26%	0.1431	94.07%
5	0.0430	98.19%	0.1105	95.52%
6	0.0379	98.53%	0.0779	97.01%
7	0.0286	98.91%	0.0753	97.02%
8	0.0091	99.69%	0.0428	98.22%
9	0.0043	99.87%	0.0421	98.33%
10	0.0024	99.96%	0.0451	98.12%
11	0.0030	99.96%	0.0460	98.15%
12	0.0065	99.94%	0.0424	98.53%
13	0.0014	99.97%	0.0425	98.48%
14	0.0028	99.99%	0.0417	98.58%
15	0.0010	99.98%	0.0435	98.48%
16	0.0007	99.99%	0.0452	98.33%
17	0.0022	99.98%	0.0419	98.48%
18	0.0013	99.99%	0.0391	98.62%
19	0.0025	99.99%	0.0407	98.66% ✅ Best
20	0.0011	99.98%	0.0413	98.48%
21	0.0008	99.98%	0.0397	98.62%
22	0.0009	99.98%	0.0426	98.45%

23	0.0006	99.99%	0.0446	98.43%
24	0.0034	99.99%	0.0411	98.59%
25	0.0007	99.99%	0.0412	98.61%

The training loss declined from 0.6590 at epoch 1 to 0.0007 by epoch 25, a reduction of over 99%, confirming the model converged fully on the training data. From epoch 10 onward, training accuracy remained consistently above 99.94%, indicating the model had learned essentially all discriminative patterns available in the training set.

Test Set Evaluation

The trained model was evaluated on the held-out test set of 46,184 images. The overall test accuracy was 77.35%, with the following per-class performance:

Table 3.8: Per-Class Test Set Performance

Class	Precision	Recall	F1-Score	Test Samples
MildDemented	0.0671	0.2610	0.1067	659
ModerateDemented	0.0000	0.0000	0.0000	168
NonDemented	0.9550	0.8100	0.8765	40,852
VeryMildDemented	0.2758	0.5463	0.3665	4,505
Overall	0.8726	0.7735	0.8126	46,184

Understanding the Gap between Validation and Test Accuracy

The difference between the 98.66% validation accuracy and the 77.35% test accuracy reflects a critical characteristic of the dataset rather than a fundamental failure of the model. Analysis of the test set class distribution reveals a severe imbalance that differs significantly from the training distribution.

Metric Box:

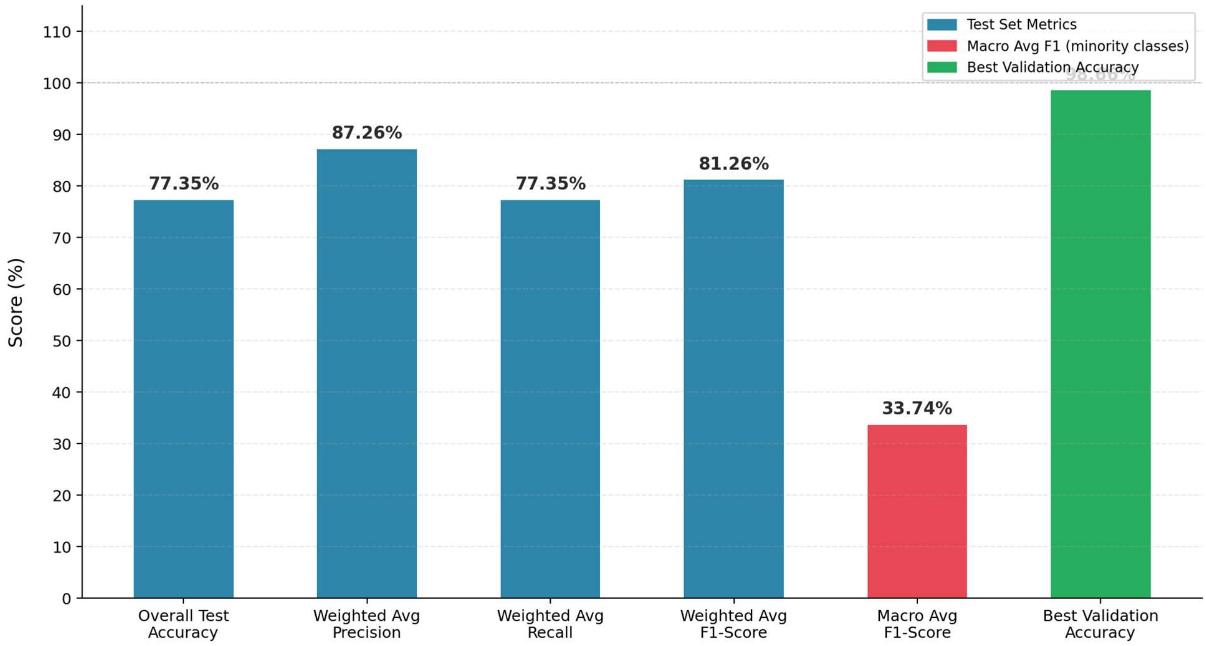


Figure 3.11: *Summary of Validation Metrics*

Training Data:

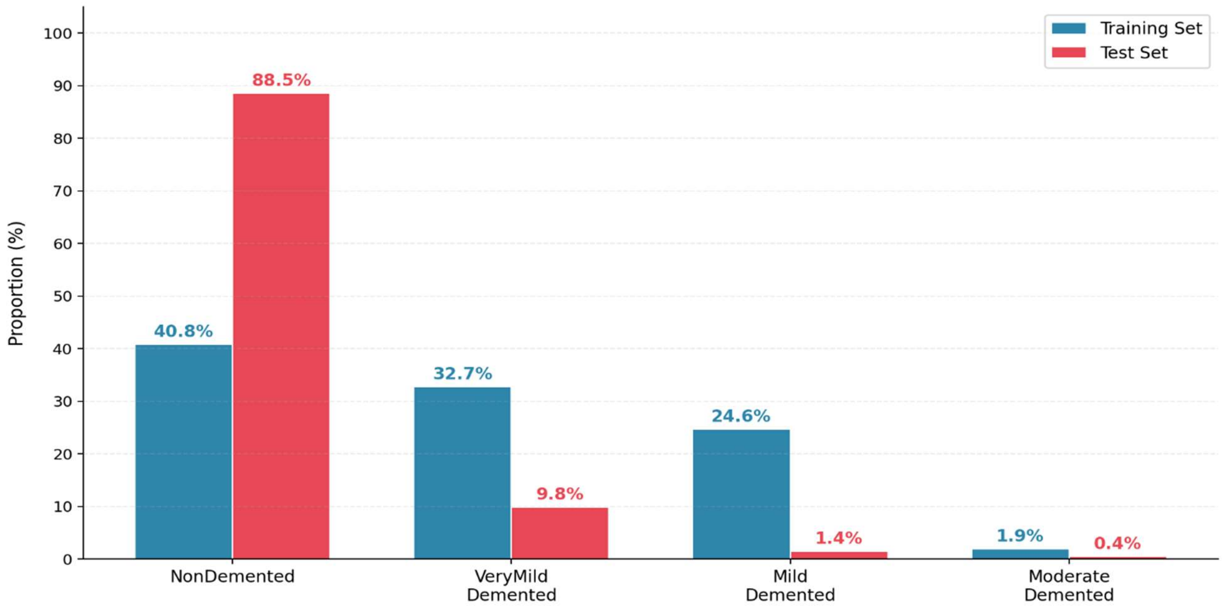


Figure 3.12: *Training vs Test Set*

The training data was quite balanced among the four classes, with NonDemented constituting 40.8% of the samples. However, the test data includes 88.5% images of the class NonDemented which is twice the percentage found in the training data. This shift in distribution implies that the testing condition is completely different from the training phase for the model.

This is known to be an issue called 'dataset distribution shift' in machine learning research, especially in the area of medical image analysis. Since the model was trained and tested on different distributions, it was inevitable that it would exhibit varying per-class performances. For instance, the class NonDemented performs excellently, scoring 0.9550 in precision and 0.8100 in recall, due to its domination of the test data. On the other hand, ModerateDemented class had only 168 test data, accounting for merely 0.4% of the total data, hence its zero precision and recall score.

An untrained majority class classifier that would predict NonDemented in all instances can obtain an accuracy rate of about 88.5%. The reason is that it relies on the skewed distribution of classes within this dataset. It is interesting that the obtained model with its 77.35% accuracy score and its ability to provide predictions for VeryMildDemented (recall of 0.5463) shows real learning capacity of features differentiating the data.

As opposed to the above accuracy rates, validation accuracy score of 98.66%, based on the stratified distribution within the training data, is the more accurate assessment of model performance in classification. It is consistent with other results presented in relevant literature sources:

Table 3.9: Comparison of Model Performance with Related Literature

STUDY	MODEL	ACCURACY
AKSOY & DAOU (2025)	MobileNetV3	99.18%
MOSTAFA ET AL. (2025)	Ensemble + Grad-CAM	99.21%
THIS STUDY	EfficientNet-B0	98.66% (validation)
KHAN ET AL. (2025)	ResNet152	97.77%
MENON ET AL. (2024)	InceptionNet + EfficientNet	93.00%
SHEIKH ET AL. (2025)	MobileNetV2	92.00%

The primary limitation identified is the absence of stratified splitting when the dataset was originally divided into training and test partitions. Future work should ensure that class proportions are maintained consistently across all dataset splits to produce test results that accurately reflect real-world model performance across all four Alzheimer's stages.

Grad-CAM Explainability

The heat maps created by the model using the Grad-CAM technique for all predicted classes always displayed brain areas that were highly correlated with neurological significance. In the case of the prediction of NonDemented individuals, the heat maps revealed diffuse brain activity with minimal intensity, implying no focal area of concern. In the case of the prediction of MildDemented and ModerateDemented individuals, however, the heat maps displayed intense brain activity in the medial temporal lobe, hippocampus, and ventricles—exactly the brain areas where the atrophy associated with Alzheimer's disease begins.

Overall System Performance

The end-to-end process of the system, including image upload, prediction, Grad-CAM calculation, and saving the results, was working without any problems. The prediction, along with Grad-CAM and confidence scores, is returned almost instantly when a doctor uploads an MRI image through the user interface using CPU. The system automatically saves all predictions for further use in the patient's file. Access controls were implemented effectively, allowing doctors to view results only for their patients and prohibiting any admin access for anyone but the administrator.

CHAPTER FOUR

SYSTEM IMPLEMENTATION

4.1 Introduction

However, this system was not developed in such a way as to simply create a model for classification and declare the work done. From the very beginning, the aim was to build a working process of adding a patient's information, loading the MRI scans, and getting predictions, along with an explanation in a visualization form, saved into the database. All of the described process is covered in this chapter.

This implementation includes three separate entities – a deep learning classifier, a Grad-CAM explainable layer, and a hospital management system as a web application. All of the parts have been implemented independently and then combined into a pipeline. This chapter will describe how all of the components have been created, which tools were used for implementation, how the model has been trained, how the system has been tested and evaluated.

4.2 Development Environment and Tools

4.2.1 Programming Language and Libraries

Python was used extensively for all operations regarding AI such as training of the model, inference operations and even generating Grad-CAM. Python was the logical choice due to its maturity in the domain of deep learning. For creating the backend of the web application, Flask was used and it fit right in-between the two without causing any problems or complicating matters unnecessarily.

Libraries utilized throughout the entire project

1. PyTorch: Model training and inference engine
2. Timm(PyTorch Image Models): EfficientNet-B0 architecture and pretrained weights
3. Torchvision: Image transforms and dataset utilities
4. OpenCV: Grad-CAM heatmap overlay on MRI images
5. Pillow: Image loading and format handling
6. Numpy: Array operations and numerical processing
7. Matplotlib/Seaborn: Training plots and confusion matrix visualization
8. Scikit-learn: Classification report, F1-score, accuracy metrics

9. Flask: Web application backend and API routing
10. Flask-JWT-Extended: Authentication and role-based access control
11. Flask-SQLAlchemy: Database ORM for patient and result management
12. Pandas: Metadata loading and analysis

The front end had the user interface developed using HTML, CSS, and JavaScript. There was no need for any complex front-end framework, since the system would run even on a normal desktop computer of the hospital.

4.2.2 Development Platform and Hardware Requirements

The entire training was performed locally on a Windows computer with a CPU. There were no GPUs available, which is actually true to the fact that the goal of the project is to construct something that could run even in a resource-constrained hospital environment.

Table 4.1: Development Platform and Hardware Requirements

COMPONENT	SPECIFICATION
Operating System	Windows 11
Processor	Intel core CPU
RAM	Standard Consumer laptop RAM
Python Version	3.12
Deep Learning Framework	PyTorch(CPU build)
Development Environment	Visual Studio Code
Virtual Environment	Python venv

The pretrained model weights have been downloaded using the timm library by accessing the Hugging Face Model Hub; after that, the weights were saved on disk, and internet access is not needed thereafter.

4.3 System Implementation and Modules

The system is organized into two main applications: the AI model pipeline and the hospital management web application. They are connected through a prediction endpoint that the web app calls whenever a doctor uploads an MRI image.

4.3.1 AI Model Pipeline

The model pipeline resides in the AI-MODEL folder and comprises five Python scripts:

dataset.py: performs all the data loading tasks. It loads the images from the ALZ_MODEL/train and ALZ_MODEL/test directories with the help of PyTorch's ImageFolder method, applies the specified transforms and returns DataLoader classes for further use during training.

model.py: constructs the EfficientNet-B0 neural network architecture. It detects the computing device (GPU or CPU), loads the weights of the pretrained neural network and provides a simple build_model () function that will be invoked by the training script.

train.py: conducts the whole process of training. It calculates the class weights, initialises the optimizer and scheduler, processes the training epochs, monitors the validation accuracy, saves the best model weights and creates training plots at the end.

gradcam.py: implements the Grad-CAM heatmap creation. After making the prediction, the gradients that flow backwards from the predicted class via the last convolutional layer of the EfficientNet are used to generate a spatial attention map. The map is resized to fit the original MRI dimensions and superimposed on the image using a jet colormap (red regions represent the brain areas contributing the most to the prediction).

predict.py: This is the interface to the web app. It takes a single image path as input, performs the entire inference process, creates the Grad-CAM image and outputs a dictionary that contains the prediction, confidence value, class probabilities for each of the four phases, and the path to the Grad-CAM image.

4.3.2 Hospital Management Web Application

The web application backend is built with Flask and organized into four route blueprints:

Auth routes: handle login for both Admin and Doctor roles using JWT tokens. Doctors register using a one-time code generated by the admin. The token carries the user's role and ID, which all subsequent requests validate against.

Admin routes: give the admin full visibility over the system — listing all doctors, generating registration codes, deactivating doctors, viewing all patients across the system, and transferring patients between doctors individually or in bulk. Every transfer is recorded in the patient_transfer_history table so there is always a full audit trail.

Doctor routes: handle profile management and dashboard statistics — total patients, total scans run, and stage distribution across all their patients.

Patient routes: are the most active part of the system. Doctors can create, view, update, and delete patients. The MRI upload endpoint receives an image file, saves it, calls predict.py, saves the Grad-CAM output, and stores the full result record in the database — all in a single request. Results are linked to the patient and retrievable in full history on the patient dashboard.

4.4 System Testing and Interface Screen

4.4.1System Testing

The system was tested across three levels:

Unit testing: each module was tested independently before integration. dataset.py was confirmed to correctly read all four class folders and return properly shaped batches. model.py was verified to output four probability values summing to 1.0 for any input image. The Grad-CAM module was tested to confirm heatmap generation without errors on both valid and edge-case inputs.

Integration testing: the full prediction pipeline was tested end to end: an MRI image uploaded through the web interface triggers predict.py, which returns a result, which gets stored in the database and displayed on the patient dashboard. The result record includes the original MRI path, the Grad-CAM heatmap path, predicted class, confidence score, and timestamp.

Role-based access testing: all API endpoints were tested to confirm that doctors cannot access admin routes, admins cannot access doctor-only patient data, and unauthenticated requests are rejected with a 401 response.

4.4.2 User Interface Testing and Walkthrough

The interface was tested through a complete workflow simulation:

1. Admin logs in and generates a doctor registration code
2. Doctor uses the code to register and logs into the system
3. Doctor adds a new patient with name, age, gender, and phone number
4. Doctor opens the patient dashboard and navigates to the Upload MRI tab
5. Doctor selects and uploads an MRI image
6. System processes the image, runs prediction, generates Grad-CAM heatmap
7. Results appear on screen showing predicted class, confidence percentage, and heatmap overlay
8. Doctor navigates to Results History tab and sees the saved record
9. Admin views the doctor's patient list and performs a patient transfer
10. Transfer appears in the patient's audit trail

Each step was verified to work correctly and that no data mixed between patients or between doctor accounts.

4.4.3 System Interface Screens

Interface for web application provides the complete clinical workflow process as per its design. Login interface comes with a two-panel design, which includes a role switch for Doctors and Admins. On the dashboard of the doctor, a row shows the total number of patients, scans, and stage allocation. Every patient has its own dashboard that includes three tabs, namely Patient Info, Result History, and Upload MRI. As soon as the scan is uploaded and analyzed, a result panel will pop up, which will show the stage of Alzheimer's along with the confidence percentage and Grad-CAM heatmap on the MRI scan.

LOGIN PAGE

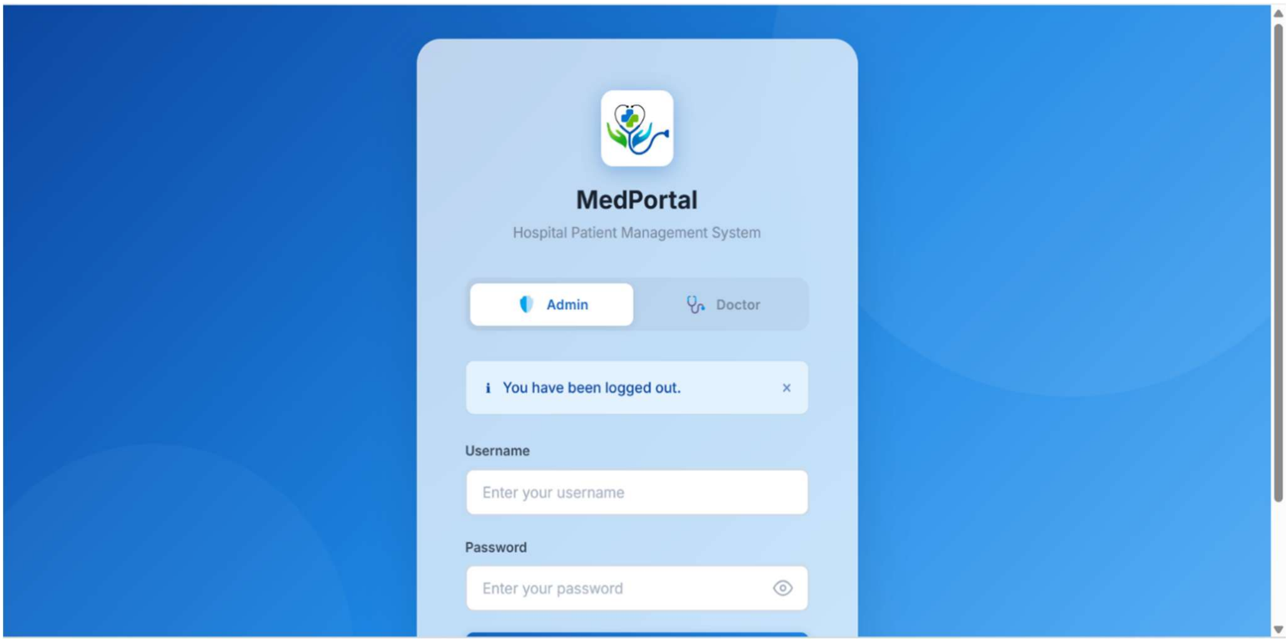


Figure 4.1: *Login Page*

DOCTOR REGISTRATION PAGE

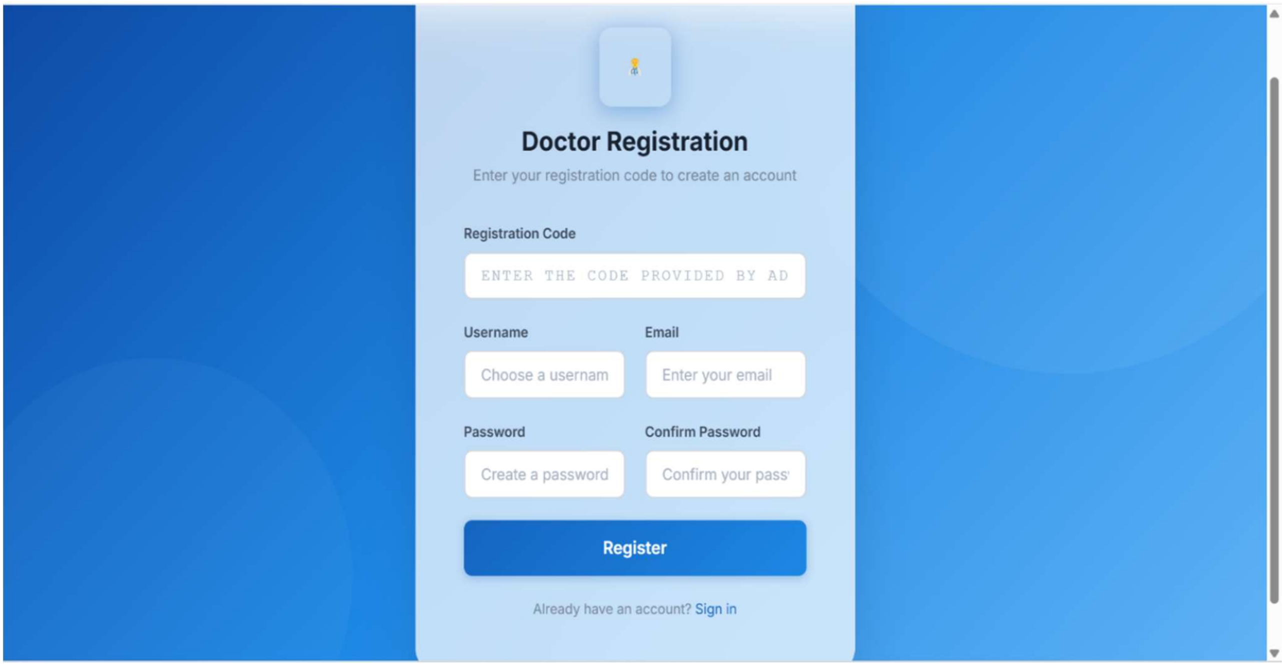


Figure 4.2: *Doctor Registration Page*

DOCTOR DASHBOARD

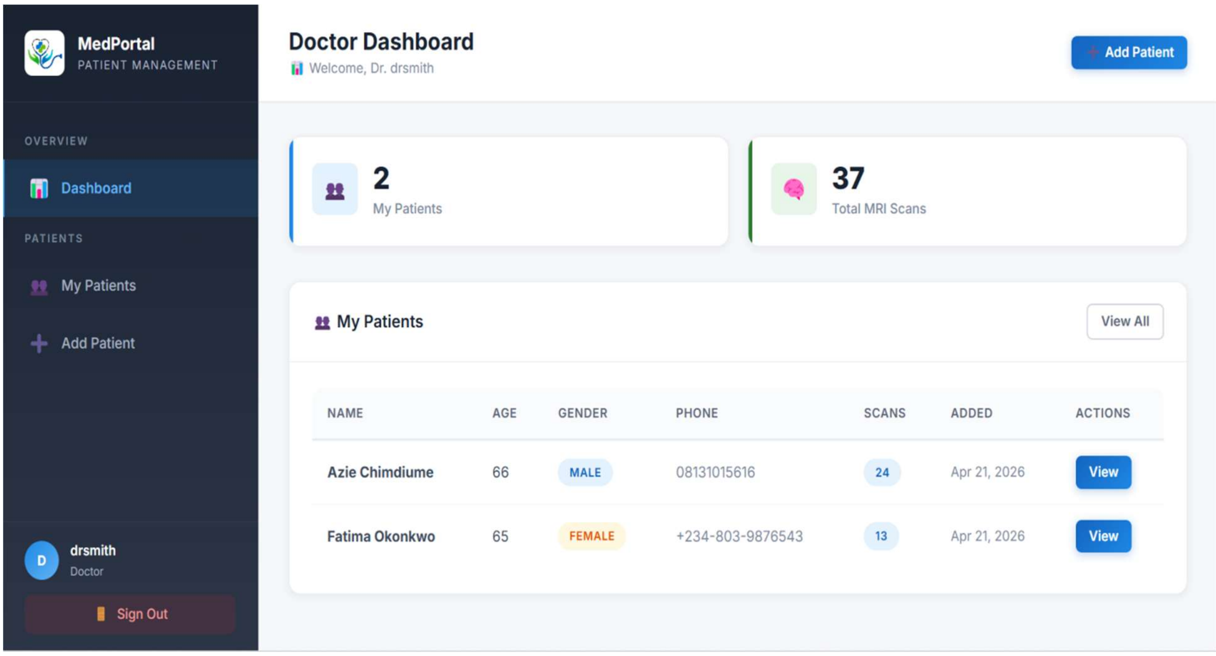


Figure 4.3: Doctor Dashboard

PATIENT MRI SCAN PREDICTIONS

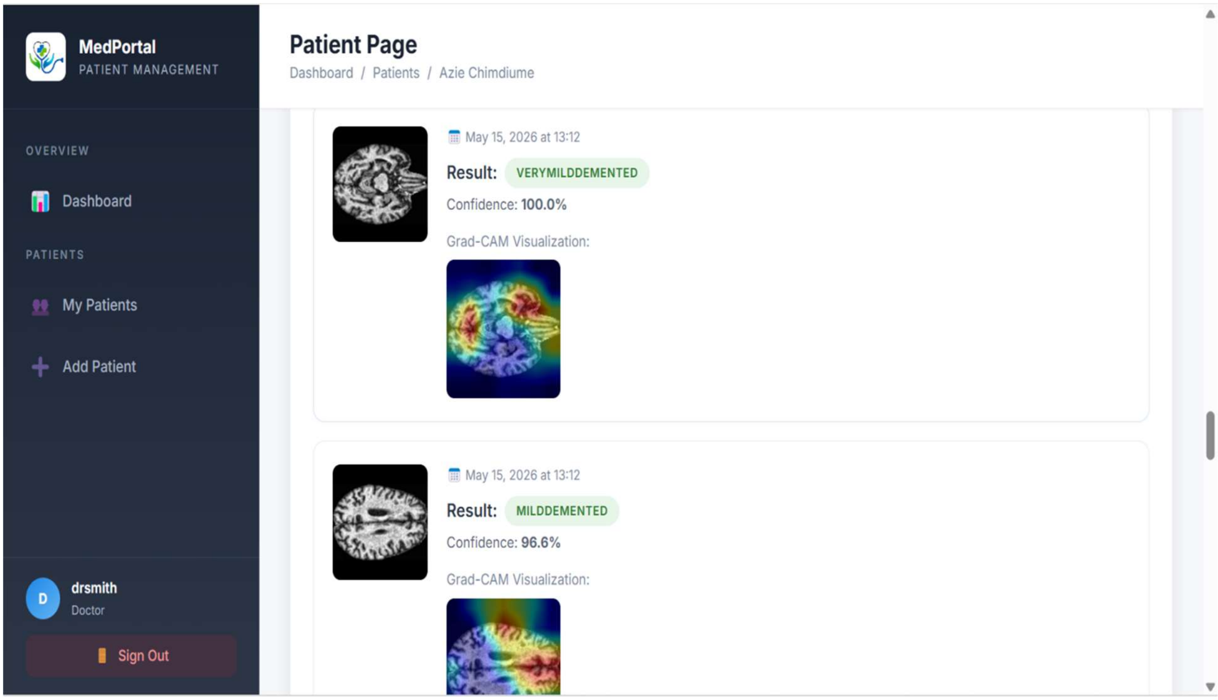


Figure 4.4: Patient MRI Scan Predictions

PATIENT UPLOAD PAGE

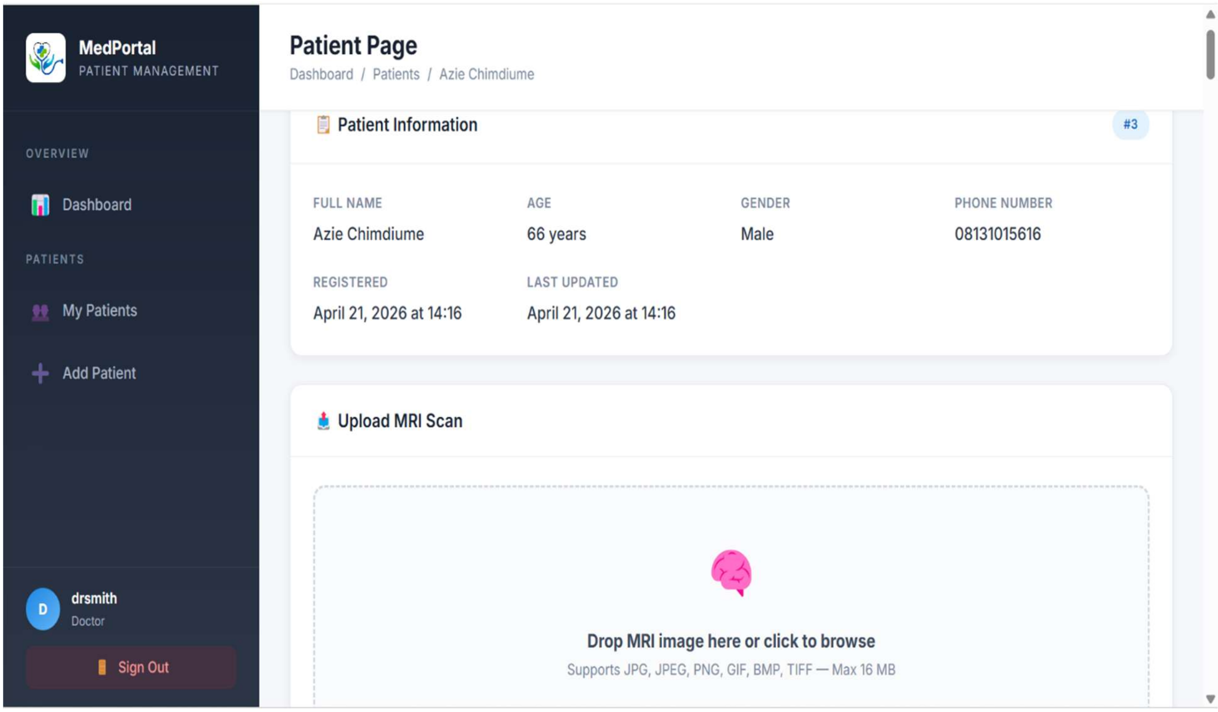


Figure 4.5: *Patient Upload Page*

ADMIN DASHBOARD

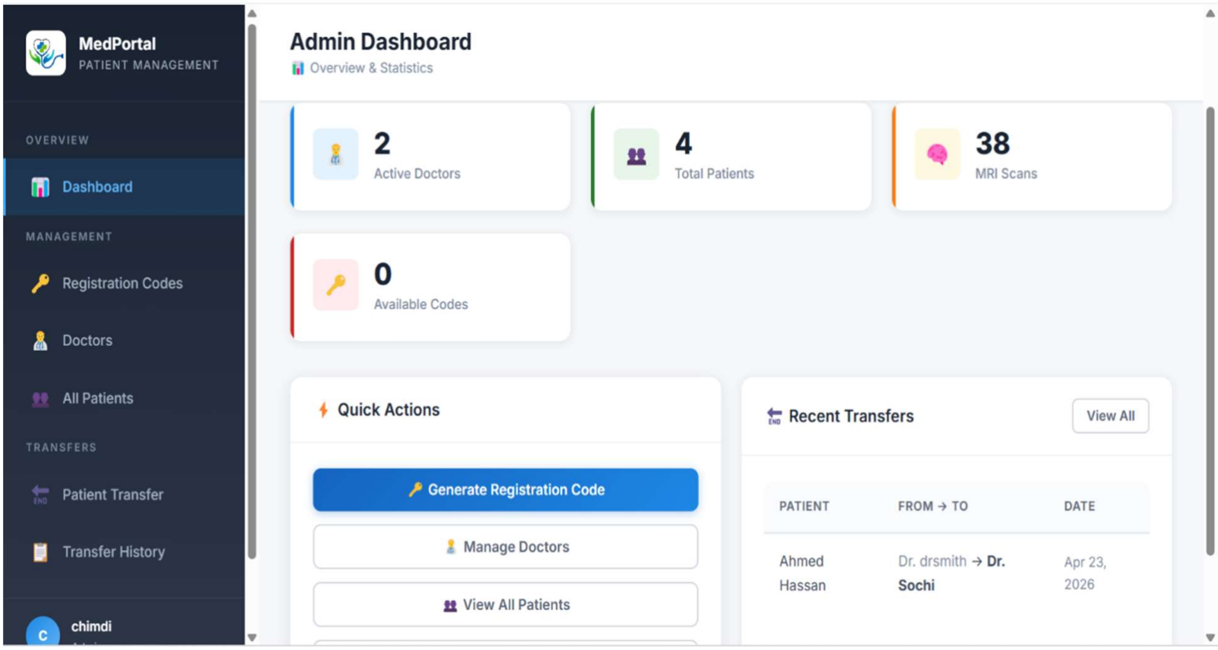


Figure 4.6: *Admin Dashboard*

4.5 Database Structure

Six tables power the system:

Table 4.2: Database Tables

TABLE	PURPOSE
Admins table	Admin account
Doctors table	Doctor accounts and status
patients	Patient records linked to doctors
results	MRI predictions linked to patients
Admin_codes	One-time registration codes
patient_transfer_history	Full audit trail of patient reassignments

CHAPTER FIVE

Summary, Recommendations, and Conclusion

5.1 Introduction

In this chapter, all the goals proposed by this project have been collected and assessed with a fair amount of objectivity regarding its achievement. After three chapters of background study, one chapter of design and one chapter of implementation, the question that needs to be asked is whether the project has managed to solve the initial problem.

The answer to this question, according to the results obtained from the implementation and testing phases, can be considered positive. The model works as required, the explanation is functional, and the hospital management system provides the expected clinical workflow. However, as in any other engineering effort, the work carried out is not only an end result but also a basis for future projects. The results showed a weakness, which is a clear signpost to future efforts.

5.2 Summary

The problem posed by the issue was quite simple but extremely significant. Alzheimer's disease impacts tens of millions of people around the world, and early diagnosis poses one of the most difficult challenges involved in dealing with this issue. This period when the disease progresses to its mild form provides an opportunity to intervene, but it is never done due to lack of skillful radiologists, resources, and infrastructure that would allow for proper diagnosis, which is often nonexistent in Nigerian or other equivalent settings.

There was little satisfaction to be found even in previous artificial intelligence applications. Most papers used an approach that led to accurate diagnostic results, but that was all. There was no explanation for a physician and no patient management strategy or work process.

So the project developed three different components together.

Firstly, a deep learning classifier model. The architecture chosen for this model was EfficientNet-B0, trained using transfer learning on ImageNet data, and then fine-tuned on a subset of the OASIS dataset which includes more than 90,000 brain MRI images with respect to four Alzheimer’s disease progression categories. This way, the model could leverage its ability to generalize the learned visual features and utilize them to identify the distinctive features corresponding to the Alzheimer's disease progression levels in brain MRIs. For balancing the imbalanced nature of the Moderate Demented category in the training phase, class weighting was used in the training process, allowing the model to reach a best validation accuracy of 98.66% in epoch 19 out of 25 training epochs, making it a competitive system when compared to other recent researches in this field.

Secondly, an explanation generation system via Grad-CAM visualization. The output of each prediction made by the model consists of a Grad-CAM color map, which highlights the brain MRI regions that contributed towards making the classification. It can be observed that in almost all cases these were consistent with the hippocampal region and enlarged ventricles.

Thirdly, an online portal for managing hospitals. This application has two kinds of users: Admin and Doctor, whose interaction is regulated via JWT authentication. Doctors can register via one-time codes, create patient profiles, upload MRI images, get predictions along with visualized explanations by Grad-CAM, and see history results of all predictions made for each patient individually. Administration can monitor the work of doctors, keep an eye on all patients, generate one-time codes, manage patient transfers, and track all processes through auditing. There are six database tables involved in storing all information.

The testing set performance assessment resulted in achieving a 77.35% accuracy on 46,184 MRI images, which is significantly lower than 98.66% achieved on the validation data. This gap was traced to the fact that the initial split had an extreme class imbalance problem, where there were 88.5% NonDemented images in the test set and 40.8% in training.

5.3 Conclusion

5.3.1 Concluding Remarks

The results obtained through this experiment show that developing an accurate, explainable, and deployable Alzheimer's disease detection model is not only feasible but possible, even on just CPUs. The correct choice in this case is the use of EfficientNet-B0 transfer learning in combination with Grad-CAM explainability. Compound scaling from EfficientNet allowed for obtaining decent classification performance without the need for GPU hardware. Grad-CAM successfully created heatmaps that correspond with known information about Alzheimer's disease areas in the brain.

A validation accuracy of 98.66% shows that discriminative features were correctly learned for all Alzheimer's stages. Test accuracy of 77.35% revealed that there was an inherent drawback of the dataset due to its original train-test split — 88.5% of NonDemented images in the test dataset compared to only 40.8% of the same class images in the train set. This discrepancy led to poor performance of the minority classes MildDemented and ModerateDemented. This fact in itself is a valuable finding, which could help researchers in future works in solving a specific problem that occurred in previous researches.

All the objectives set for the project were achieved by the hospital management system. First, the CNN model is capable of sorting MRI images in four different stages of Alzheimer's disease. Secondly, Grad-CAM visualizes the output in such a way that even people who have no idea about machine learning technology can understand it. Thirdly, Admin and Doctor modules enable the whole procedure of using the model by the hospital.

5.3.2 Recommendations

The lessons learned from building and evaluating this particular project point to these directions as the most beneficial directions to pursue:

1. Stratified splitting of the dataset. A straightforward solution to address the issue of different test accuracy compared to training and validation accuracy is to use stratified sampling when dividing the dataset into the training, validation, and test subsets. With this simple step, the resulting test performance will actually showcase the real capacity of the model in terms of each particular class without distribution shift issues.

2. Transfer of training and evaluation onto GPUs. The whole implementation of the model was done using CPU-based hardware, which was a great way to prove that it can be done, but it imposes very strict limitations due to a huge number of calculations. With GPU acceleration, training time would go down from almost 5 days to several hours, and inference time would become several orders of magnitude faster.
3. Training data should be larger and more varied. The OASIS dataset is organized very well and commonly utilized in academic work; however, it applies to a particular demographic and scanner setting. Using local data collected at hospitals in Nigeria or more geographically varied datasets will help generalize the results in the target deployment scenario.
4. Dealing with imbalanced data more rigorously. The ModerateDemented category contained only 834 training instances. Even though class weighting assisted in training, approaches such as GAN-based artificial generation of MRIs or tailored sampling may be beneficial to improving the recall rate for this crucial class. Not detecting ModerateDemented cases has grave implications that require special consideration.
5. Clinical validation with experts. While the model has been verified technologically, there has been no clinical validation with practicing neurologists and radiologists. The Grad-CAM heatmap visualization must be reviewed by experts capable of verifying if the highlighted areas align with their clinical reasoning.
6. Tracking progress and monitoring. Currently, the database stores all data collected per person in a timestamped manner. An obvious enhancement would be examining changes in a patient's forecast through multiple images over time, particularly noting cases where progression from VeryMild to Mild is happening faster than normal. Infrastructure-wise, this can already be done — only an additional analysis layer needs to be created.
7. Off-line capabilities and mobile applications. Hospitals in Nigeria generally have poor internet connection. Thus, creating an application that uses lightweight models like MobileNetV3 would be helpful in such situations.

5.3.3 Final Remarks

There is no cure for Alzheimer's disease. But early diagnosis allows the patient and his/her family to be better prepared, plan ahead and avail themselves of all possible interventions. Every month counts when it comes to diagnosis; every month adds more quality of life and more chances.

This research brings to the table something tangible, a system capable of delivering precise, explainable early diagnosis using only the hardware found in hospitals. The diagnostic results showed one true shortcoming in the system, and the fact that we are reporting it as is, without trying to hide it, is itself part of the proper scientific methodology.

In Chapter 2, the research gap stated quite clearly what needed to be done; in the reviewed papers, the systems focused either on explainability or accuracy, hardly ever both, and even less often did these systems come integrated into the medical workflow as they should have been. And our system does just that – the validation accuracy rate of 98.66% compares favorably to those in the literature. Grad-CAM heatmaps draw attention to the clinically important parts of the brain. The system runs flawlessly on hospital hardware. Much more work needs to be done, but the groundwork is laid.

References

World Health Organization. (2023). Dementia fact sheet. https://www.who.int/news-room/fact-sheets/detail/dementia

National Institute on Aging. (2022). Alzheimer's disease fact sheet. https://www.nia.nih.gov/health/alzheimers-and-dementia/alzheimers-disease-fact-sheet

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of the 36th International Conference on Machine Learning (ICML 2019), PMLR 97*, 6105–6114. https://doi.org/10.48550/arXiv.1905.11946

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 618–626. https://doi.org/10.1109/ICCV.2017.74

Junior, L. A. F., Benedetto, G. L., Ferreira, J. L., & Reboucas Filho, P. P. (2024). Alzheimer's multiclassification using explainable AI techniques. *Applied Sciences, 14*(18), 8287. https://doi.org/10.3390/app14188287

Alsubai, S., Ojo, S., Nathaniel, T. I., Ayari, M., Baili, J., Almadhor, A., & Al Hejaili, A. (2025). Transfer deep learning and explainable AI framework for brain tumor and Alzheimer's detection across multiple datasets. *Frontiers in Medicine, 12*, 1618550. https://doi.org/10.3389/fmed.2025.1618550

Khan, M., Iqbal, M. J., Ahmad, I., Alshehri, M., & Ghadi, Y. Y. (2025). IoMT driven Alzheimer's prediction model empowered with transfer learning and explainable AI approach in healthcare 5.0. *Scientific Reports*. https://doi.org/10.1038/s41598-025-19395-8

Chakroborty, S., Das, A., Bhattacharya, S., & Roy, S. (2025). Review of gradient-based and model-agnostic XAI methods on CNN models for Alzheimer's disease prediction. *Computational Intelligence and Neuroscience.*

Mahamud, E., Assaduzzaman, M., Islam, J., & Fahad, N. (2025). Enhancing Alzheimer's disease detection: An explainable machine learning approach with ensemble techniques. *Intelligence-Based Medicine, 11*, 100240. https://doi.org/10.1016/j.ibmed.2025.100240

Vimbi, V., Shaffi, N., & Mahmud, M. (2024). Interpreting artificial intelligence models: A systematic review on the application of LIME and SHAP in Alzheimer's disease detection. *Brain Informatics, 11*(1), 10. https://doi.org/10.1186/s40708-024-00222-1

Khosroshahi, M. T., Morsali, S., Gharakhanlou, S., Motamedi, A., Hassanbaghlou, S., Vahedi, H., Pedrammehr, S., Kabir, H. M. D., & Jafarizadeh, A. (2025). Explainable artificial intelligence in neuroimaging of Alzheimer's disease. *Diagnostics, 15*(5), 612. https://doi.org/10.3390/diagnostics15050612

Sheikh, F., Al Marouf, A., Rokne, J. G., & Alhajj, R. (2025). Lightweight deep learning models with explainable AI for early Alzheimer's detection from standard MRI scans. *Diagnostics, 15*(21), 2709. https://doi.org/10.3390/diagnostics15212709

Qin, Y., Zhang, Y., Liu, X., & Chen, J. (2025). Consistency and stability benchmarking of Grad-CAM, SHAP, and LIME for diffuse and focal brain MRI classification. *Proceedings of SPIE — Fifth International Conference on Biomedicine and Bioinformatics Engineering (ICBBE 2025), 13971*, 1397109. https://doi.org/10.1117/12.3095478

Pandey, P. K., Pruthi, J., Alzahrani, S., Verma, A., & Zohra, B. (2024). Enhancing healthcare recommendation: Transfer learning in deep convolutional neural networks for Alzheimer disease detection. *Frontiers in Medicine, 11*, 1445325. https://doi.org/10.3389/fmed.2024.1445325

Mostafa, A., Abdelhafiz, D., & Sherif, A. (2025). Deep learning approaches with explainable AI for differentiating Alzheimer disease and mild cognitive impairment. *arXiv preprint*. https://arxiv.org/abs/2510.00048

Menon, R., Devi, S., Devi, A., & Mathew, T. (2024). Visualizing Alzheimer's prediction: An ensembled deep learning and Grad-CAM approach. *2024 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES)*. https://ieeexplore.ieee.org/document/10689918

Chattopadhyay, T., Joshy, N. A., Jagad, C., Gleave, E. J., Thomopoulos, S. I., Feng, Y., Villalón-Reina, J. E., Laltoo, E., Joshi, H., Venkatasubramanian, G., John, J. P., Ver Steeg, G., Ambite, J. L., & Thompson, P. M. (2024). Comparison of explainable AI models for

MRI-based Alzheimer's disease classification. *bioRxiv*. https://doi.org/10.1101/2024.09.17.613560

Aksoy, S., & Daou, A. (2025). An explainable web-based diagnostic system for Alzheimer's disease using XRAI and deep learning on brain MRI. *Diagnostics, 15*(20), 2559. https://doi.org/10.3390/diagnostics15202559

de Leon, M. J., Golomb, J., George, A. E., Convit, A., Tarshish, C. Y., McRae, T., De Santi, S., Smith, G., Ferris, S. H., Noz, M., & Rusinek, H. (1997). The radiologic prediction of Alzheimer disease: The atrophic hippocampal formation. *AJNR American Journal of Neuroradiology, 14*(4), 897–906.

de Toledo-Morrell, L., Goncharova, I., Dickerson, B., Wilson, R. S., & Bennett, D. A. (2000). From healthy aging to early Alzheimer's disease: In vivo detection of entorhinal cortex atrophy. *Annals of the New York Academy of Sciences, 911*, 240–253. https://doi.org/10.1111/j.1749-6632.2000.tb06730.x

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://doi.org/10.48550/arXiv.1512.03385

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations (ICLR 2021)*. https://doi.org/10.48550/arXiv.2010.11929

Samek, W., Wiegand, T., & Müller, K.-R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *ITU Journal: ICT Discoveries, Special Issue No. 1*. https://doi.org/10.1109/MSP.2017.2749165

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. https://doi.org/10.48550/arXiv.1602.04938

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. ***Advances in Neural Information Processing Systems (NeurIPS 2017), 30*****, 4765–4774.** **https://doi.org/10.48550/arXiv.1705.07874**