Universidad de Valladolid

PROGRAMA DE DOCTORADO EN:

TECNOLOGÍAS DE LA INFORMACIÓN Y LAS TELECOMUNICACIONES

-

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE TELECOMUNICACIÓN

DPTO. DE TEORÍA DE LA SEÑAL Y COMUNICACIONES E INGENIERÍA TELEMÁTICA

TESIS DOCTORAL:

DEEP-AD: The Deep Learning Model for Diagnostic Classification and Prognostic Prediction of Alzheimer's disease

Presentada por

D. Deevyankar Agarwal

Para optar al grado de Doctor por la Universidad de Valladolid Dirigida por

Prof. Isabel De la Torre Díez

Dr. M. Alvaro Berbís

Valladolid, 2023

Abstract

Approximately 50 million individuals globally are suffering from Alzheimer's disease (AD) and other forms of dementia, which not only affects them but also has an impact on their families and caregivers. As a result, there is a growing interest in discovering new methods for accurately predicting the progression of AD. **Early detection of AD and its prodromal stages, such as mild cognitive impairment (MCI), is crucial for a possible illness delay**. One of the numerous biological indicators of the illness is known to be sensitive to structural abnormalities of the brain, which may be shown on neuroimaging biomarkers such as magnetic resonance imaging (MRI) scans. Automatic categorization of AD may be possible with machine learning (ML) models by using MRI scans, and with deep learning (DL) approaches that can extract features from high-dimensional data like MRI.

In terms of context, the **aim of this dissertation is to aid neuroradiologists in their clinical judgment regarding the early detection of AD by using DL**. To that aim, the system design research methodology is suggested in this dissertation for achieving **three goals**.

The **first goal** is to investigate the DL models that have performed well at identifying patterns associated with AD, as well as the accuracy so far attained, limitations, and gaps. **A systematic review of the literature (SLR)** revealed a shortage of empirical studies on the early identification of AD through DL. In this regard, **thirteen empirical studies were identified and examined**. **We concluded** that three-dimensional (3D) DL models have been generated far less often and that their performance is also inadequate to qualify them for clinical trials.

The **second goal** is to provide the neuroradiologist with the computer-interpretable information they need to analyze neuroimaging biomarkers. Given this context, the next step in this dissertation is to **find the optimum DL model to analyze neuroimaging biomarkers**. It has been achieved in **two steps**. In the **first step**, eight state-of-the-art DL models have been implemented by training from scratch using end-to-end learning (E2EL) for two binary classification tasks (AD vs. CN and AD vs. stable MCI) and compared by utilizing MRI scans from the publicly accessible datasets of neuroimaging biomarkers. Comparative analysis is carried out by utilizing efficiency-effects graphs, comprehensive indicators, and ranking mechanisms. For the **training of the AD vs. sMCI task, the EfficientNet-B0 model** gets the highest value for the comprehensive indicator and has the fewest parameters. **DenseNet264 performed** better than the

others in terms of evaluation matrices, but since it has the most parameters, it costs more to train. For the **AD vs. CN task by DenseNet264**, we achieved **100% accuracy for training and 99.56% accuracy for testing**. However, the classification accuracy was still **only 82.5% for the AD vs. sMCI task**. In the **second step, fusion of transfer learning (TL) with E2EL** is applied to train the **EfficientNet-B0 for the AD vs. sMCI** task, which achieved **95.29% accuracy for training and 93.10% accuracy for testing**. Additionally, we have also **implemented EfficientNet-B0 for the multiclass AD vs. CN vs. sMCI classification task with E2EL to be used in ensemble of models** and achieved 85.66% training accuracy and 87.38% testing accuracy.

To evaluate the model's robustness, neuroradiologists must validate the implemented model. As a result, the **third goal of this dissertation is to create a tool that neuroradiologists may use at their convenience**. To achieve this objective, this dissertation **proposes a web-based application (DEEP-AD)** that has been created by making an **ensemble of Efficient-Net B0 and DenseNet 264 (based on the contribution of goal 2)**.

The accuracy of a DEEP-AD prototype has undergone repeated evaluation and improvement. First, **we validated 41 subjects of Spanish MRI datasets (acquired from HT Medica, Madrid, Spain), achieving an accuracy of 82.90%, which was later verified by neuroradiologists**. The results of these evaluation studies showed the accomplishment of such goals and relevant directions for future research in applied DL for the early detection of AD in clinical settings.

Keywords : Alzheimer's; Deep Learning; End-to-End Learning; Transfer Learning; Ensemble Learning; Convolutional Neural Network; Mild Cognitive Impairment; MRI; EfficientNet; DenseNet; Spanish MRI Dataset.

Acknowledgments

This dissertation represents the finish line of a lengthy journey, during which I was closely accompanied by individuals who were always there for me, who helped me discover the best route forward, and who turned my exhaustion into drive. The following paragraphs are dedicated to all of them.

First, I would like to start thanking my supervisors, Professor Isabel De la Torre Díez and Dr.Álvaro Berbís Moreno . Their undying commitment and wise counsel during this time served as a solid foundation upon which I could build my confidence to make the best possible choices and to never feel abandoned. As a result of their hard work, this dissertation is now a reality, and I owe them a great debt of gratitude. I would also like to express my gratitude to Dr. Alvaro for providing the dataset of MRI scans of the Spanish subjects from HT Medica, Spain, which was used to validate DEEP-AD.

At this point, I can't stop thinking about Professor Yannis Dimitriadis, who gave me the foundational knowledge I needed to begin my research career. His encouragement and insight throughout the process have been invaluable resources.

I would also like to express my gratitude to the rest of the GTe-UVA research group for coming along for the ride. I feel very fortunate to be a part of this community, where the emphasis is placed on fostering positive relationships and developing one's own character. Thanks especially to Susel Gongora Alonso, who helped guide my study in its formative years. Prof. Gonçalo Marques is much appreciated for coming to my research sessions and making useful contributions that helped me move my dissertation forward. Sincere appreciation goes to everyone who helped create the DEEP-AD.

Without the help of friends and family and the downtime spent with them, I would not have been able to complete my dissertation. My wife Aparna and son Jai, who have been there for me through good and bad times alike, are among the people I would want to thank.

Finally, I will never be able to repay my parents for all the love and understanding they've shown me throughout my life. This may seem like generic appreciation for loved ones, but focusing on my dissertation has helped me appreciate the tremendous help my family has given me on a regular basis. They will support you no matter what life throws at you.

With all one’s heart, thank you!

Deevyankar Agarwal

Contents

Table of Figures

List of Tables

List of Algorithms

Chapter 1: Introduction

Summary: This chapter outlines the overall research context of the dissertation, the research question, the major goals, and the approach used to achieve them. The dissertation is about using techniques based on **machine learning (ML)** and **neuroimaging biomarkers** to find early signs of **Alzheimer's disease (AD)**. More specifically, we want to provide conceptual and technical tools to help professionals (such as neuroradiologists and doctors) make decisions about the diagnosis of the various stages of AD. Additionally, we want to increase the existing, sparse corpus of empirical research on the effects of various ML-based strategies for analyzing high-dimensional medical data.

We suggest **three key contributions** in line with the System Design Research Methodology to tackle the current issues with early detection of AD**: an analysis of the impact of machine learning (ML)-based strategies for diagnosing AD**, the identification and **implementation of the optimum deep learning (DL) models**, and a **web-based tool** to assist in the prediction of various stages of AD. The contributions have been assessed and improved throughout the course of the study, utilizing thirteen empirical investigations for the early diagnosis of AD using ML-based techniques as well as a series of experiments with neuroradiologists.

1.1 Motivation

The most prevalent kind of dementia is AD. It is a neurodegenerative disorder that progresses and cannot be reversed. There are no reports of any proven disease-altering treatments. Many researchers believe that AD has a slow-moving course that probably starts many years before any clinical symptoms appear [1] . By the year 2022, there were around 6.5 million Americans living with AD, and it has been estimated that by the year 2050, this number would increase to about 14 million [2]. In the USA, the cost of managing AD in 2022 was anticipated to be over USD 321 billion, including social, medical, and economic losses to patients' family members. By 2050, this expenditure might reach USD 1 trillion. Maria Luisa Carcedo, Spain's Minister of Health, Consumption, and Social Welfare, has outlined the key areas of the nation's National Alzheimer's Plan 2019–2023[1] , where she hopes to enhance AD diagnosis.

[1] https://www.alzheimer-europe.org/news

To categorize individuals who are in the early stages of their clinical development of AD, the term "mild cognitive impairment" (MCI) is used. Clinical trials focused on these patients [3]. MCI could be viewed as a stage of transition between AD and normal aging. In order to aid in the early detection of AD, it will be useful to distinguish between MCI patients who will gradually advance to AD and those who do not exhibit any clinical signs of AD. Consequently, a multi-class decision system is needed to distinguish AD and its many stages from normal controls. Since AD may be detected without the help of any specialists or skills when it is too late for therapy, it is crucial to distinguish it from MCI or normal individuals. The biggest difficulty is differentiating MCI from normal controls and predicting the development of MCI into AD in order to provide patients with prompt therapy.

A rapid advancement in neuroimaging methods, such as magnetic resonance imaging (MRI), positron emission tomography (PET), and functional MRI (fMRI), produces a large amount of high-dimensional neuroimaging data that is used by neuroradiologists and in ML approaches for the automatic early detection of AD. Competitions like the Alzheimer's disease Big Data Competition [4] and the challenge of detecting MCI using MRI data[5] also increased the interest of researchers in solving this issue. These competitions aimed to utilize an open science strategy to swiftly uncover reliable predictive biomarkers for AD that may be used by the research, pharmaceutical, and regulatory communities to advance the diagnosis and treatment of AD. Many Multiple ML methods have been developed and have shown excellent performance [6]–[11].

Several ML pattern analysis methods, including support vector machines (SVMs), logistic regression (LR), and SVM-recursive feature elimination [12], have been successful in detecting AD. Automated SVM-based diagnostic ML models for neuropsychiatric illnesses, on the other hand, need human-created features because they are unable to extract adaptive characteristics.

These approaches also need technical expertise and go through numerous phases of optimization [13], which may be time-consuming and difficult. Extraction of features, feature selection, dimension reduction, and the use of a classification algorithm are the four main phases involved. However, problems in reproducing these implementations have emerged [14].

Because of the rising power of GPUs, DL methods for image classification applications have evolved. DL is a subset of ML that models how the human brain works to identify complex patterns. It learns features, latent interpretations, and disease-related designs through spontaneous training and explore associations in different parts of images. Researchers from different fields,

such as high-dimensional medical image analysis, image segmentation, object recognition, and disease detection, have been utilizing DL models. Statistical ML methods were initially effective in automating the diagnosis of AD, but recent advances in DL methods like convolutional neural network (CNN) and sparse autoencoders [15]–[22] have outperformed them. The use of CNNs has rapidly spread into a variety of domains, beginning with AlexNet outstanding performance in ImageNet large-scale visual recognition challenge [23] and then expanded into medical image analysis, started with 2D images, such as chest X-rays [24], and then progressed onto 3D images, including MRI. End to End Learning (E2EL) approach serves as the bedrock of DL. The primary advantage of E2EL [25] is that it may potentially improve all stages of the training pipeline simultaneously.

Transfer learning (TL) is becoming more popular in the field of DL because it allows for successful DL training even when there is a lack of data [26], [27] . Human behavior, which involves learning from the past to solve difficult challenges, is the basis for the TL concept. TL aids in the acceleration of the training process and the enhancement of the performance of DL architectures. DL approaches based on CNN have shown excellent performance, especially for networks with similar training and testing parameters [28]. This is also correct in the context of AD-related research, as the majority of them have utilized the public data set of the ADNI [29] for training and testing. One limitation of DL is that it is difficult to adjust the possible bias in the network once the complexity is extremely high to assure reproducibility and transparency.

Diagnosing AD accurately and efficiently is crucial for starting appropriate therapy. Anticipatory diagnosis of AD is particularly important for the advancement of therapeutics and, eventually, for successful patient care. More recent applications of DL strive to achieve analysis speeds and accuracy that surpass those of human practitioners.

The famous Google research [30] on diabetic retinopathy diagnostics showed superior classification performance compared to that of a trained human expert. However, there are no DL-based systems deployed in clinical settings to assist experts in taking decisions about the progression of AD. For DL to be useful in diagnostics, the predicted classifier must be interpretable, and performance consistency across situations is essential. Future research must inevitably replicate major discoveries from DL on completely independent data sets. This is now acknowledged in genetics [31], but has been sluggish to enter DL research using neuroimaging

data. Hopefully, the increasing open biology of medical research data, particularly in the area of AD and associated illnesses, will offer a foundation for addressing this issue.

This chapter continues by outlining the efforts made in this dissertation to address such constraints. The dissertation's major purpose and the sub-goals under which the research issue was framed are described in Section 1.2; the dissertation's research method is outlined in Section 1.3; and the dissertation's organization and findings are summarized in Section 1.4.

1.2 Dissertation Goals and Contributions

In considering the preceding research context, this dissertation addresses the following overarching issue:

"How can neuroradiologists be better assisted in determining the stage of AD so that appropriate therapy may be started without unnecessary delay?"

In this dissertation, I want to find out how neuroradiologists can use neuroimaging biomarkers and DL-based strategies to track the progression of AD so that patients can get treatment right away. In addition, I want to understand and contribute with a series of empirical studies, the impact of DL-based tactics in real-world clinical settings and to acquire helpful insights for the creation of the conceptual and technical tools presented in the dissertation before offering such assistance. I suggest that the following three intermediate goals (**see Figure. 1-1**) be met to answer the research question posed in this dissertation:

1. **To understand the DL models that have performed well at identifying patterns associated with AD by using the neuroimaging biomarkers.**

   DL-based techniques for medical image classification tasks may effectively learn features, hidden representations, disease-related patterns, and investigate correlations in diverse areas of images via impulsive learning. Researchers have had trouble putting DL-based strategies for early AD detection into place because of the unique properties of neuroimaging biomarkers, such as high-dimensional heterogeneous data, similar brain anatomies in MCI and AD patients, and the lack of a larger number of neuroimaging biomarkers. The use and acceptance of DL-based strategies in such environments are hindered by a lack of empirical proof of their efficacy in early AD detection.

   In view of this, the **first major contribution** of this dissertation (CONT#1) is **a compilation of thirteen empirical studies analyzing the effect of the DL strategies on neuroimaging biomarkers for the early detection of AD**. These analyses aim to

understand the effect of DL-based strategies for early detection of AD and to motivate the usefulness of technological systems supporting the use of these strategies in clinical settings.

2. **To provide the neuroradiologist with the computer-interpretable information they need to analyze an MRI scan.**

As mentioned before, the number of people with MCI is growing, and it's hard for neurologists to predict how the disease will progress. This makes it necessary to use automated methods to help in taking the decision about the progression of MCI. The computer-interpretable representation of neuroimaging biomarkers is presented as a potential solution for the automatic early detection of AD. Furthermore, this digital representation would allow successful DL models to be reused in the detection of other diseases. However, the current support of DL-based strategies is limited. Most current DL models stick with the same data source ADNI, and don't utilize E2EL; therefore, they can't provide a generic solution and aren't suitable for use in clinical settings.

Therefore, one of the main contributions (CONT#2)of this dissertation is the **identification and implementation of the optimum DL model based on a novel approach, "The Fusion of E2EL and 3D TL," and MRI scans** that enable a good classification of MRI scans in various phases of AD.

3. **A tool that neuroradiologists may use at their convenience to evaluate the model's robustness.**

As it was previously mentioned, the model's responsiveness must be verified, and the deployed system must also be conveniently accessible via some online medium in clinical settings to assist neuroradiologists. Therefore, **"DEEP-AD," a web-based tool to aid neuroradiologists in their clinical judgment regarding the early detection of AD**, is another significant contribution (CONT#3) of this dissertation. DEEP-AD is founded on an ensemble of 3D CNNs(EfficientNet-B0 and DenseNet-264 architectures) with E2EL and 3DTL. Furthermore, it has been validated on 41 individuals using Spanish MRI datasets, with an 88.70% success rate. This data collection was given by HT Medica of Spain.

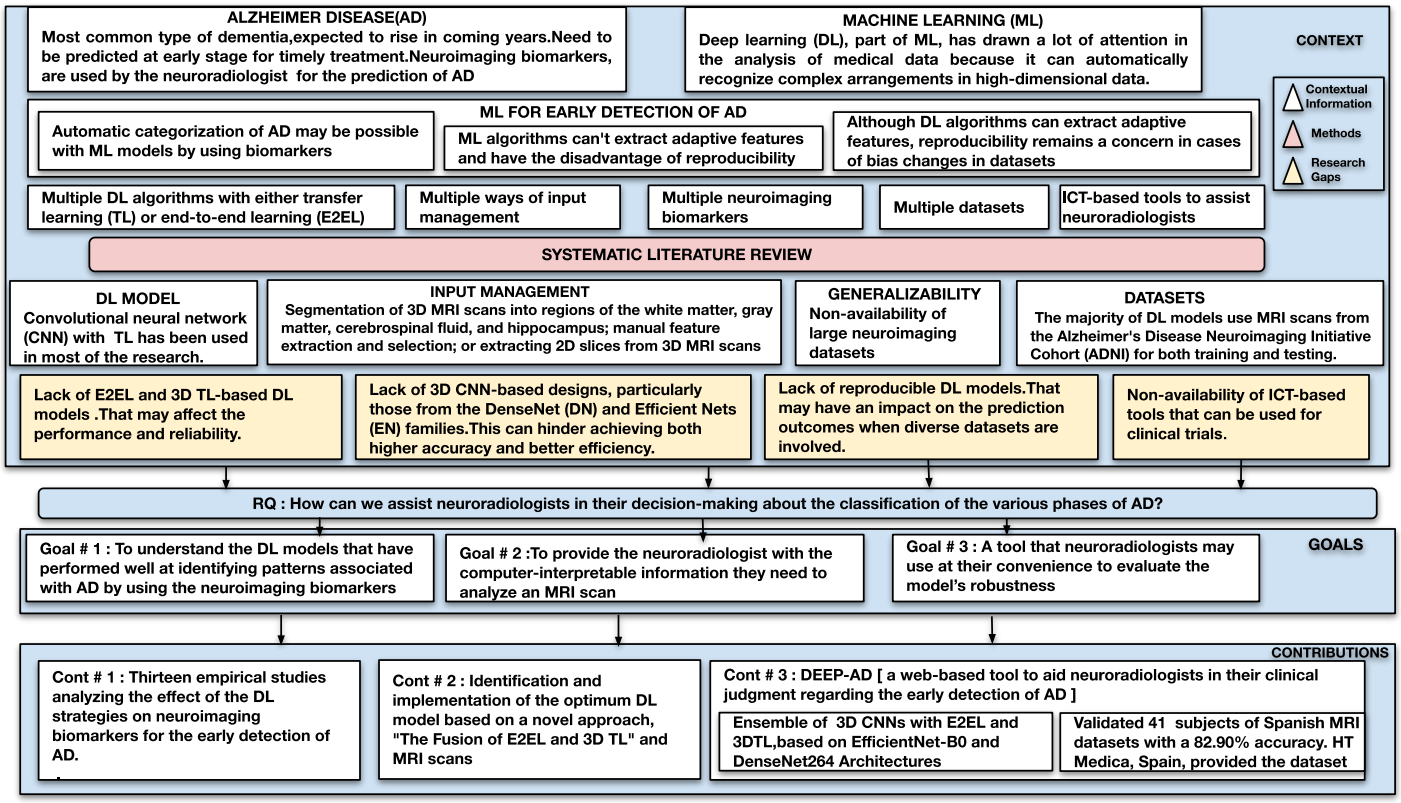


Figure 1-1 General dissertation schema including the context, research question, goals, and contributions.

1.3 Research Methodology

Choosing a suitable research approach is crucial for achieving the goals. The term "methodology" is used to describe "the assumptions and attitudes that serve as a reason for study and the standards or criteria the researcher employs to analyze evidence and reach a conclusion" [32]. The study field, the research questions and goals, and the researcher's own personal philosophy all have a role in the decision on which research technique to use. Therefore, the researchers' worldview or paradigm [33] is one of the first things they should think about before beginning the study process.

According to Creswell and Poth [34] , there are four main worldviews, which vary in how this world is imagined (i.e., ontology) and how knowledge is generated (i.e., epistemology): Social constructivism, the transformational paradigm (advocacy/participatory), post-positivism, and pragmatism.

My own perspective is more deterministic (every occurrence is determined by previously existing causes and circumstances) and hence more connected with the post-positivism worldview. According to this worldview, there is a single reality that can only be understood imperfectly and

probabilistically because of many aspects such as human environment, background, culture, genes, and so on . Furthermore, this worldview's philosophy is based on the objective creation of knowledge and the generalization of outcomes .

Since most of the variables in exact sciences like physics or chemistry are well-known and under human control, this strategy works effectively in those fields. Nevertheless, I discovered while pursuing my PhD that the social sciences face a unique set of challenges because of the sheer volume and complexity of the factors that influence human behavior. Therefore, while this reality can be quantified through probabilities based on various social factors such as background, it is also crucial to incorporate qualitative methods and subjective interpretations, such as observations and open-ended survey questions, to provide context and gain insights into individuals' emotions and perceptions, and to further comprehend the underlying causes of these probabilities.

Thus, in this dissertation, we analyze the presence of a reality that, social circumstances may be understood differently by people and the knowledge can be attained via various ways and methodologies depending on the study purpose and environment. As a result, this way of understanding the world and constructing knowledge is better aligned with the pragmatist worldview [35]. According to this viewpoint, research must be contextualized, and information is gathered using both quantitative and qualitative approaches based on the study objectives, with an emphasis on the practical implications.

As previously mentioned, this dissertation is structured in a multidisciplinary area where conceptual and technical contributions are intended to have an influence on the medical sector. As a result, we investigated many common research approaches, including the Engineering Method [36], Design Science Research Methodology [37], and System Development Research Methodology [38]. Finally, the System Development Research Methodology (SDRM) was chosen as the best methodology for this dissertation for the reasons listed below.

SDRM integrates research concerns often seen in the social (behavioral) and technical (software development) sectors, both of which are closely related to the aims of this dissertation. Also, SDRM is a good way to study because it includes system development as a major part of the study process. The overarching objectives of this dissertation, namely, the understanding of DL strategies in analyzing the neuroimaging biomarkers for early prediction of AD and providing an affordable tool to neuroradiologists through the development of technological and conceptual contributions, are completely aligned with the purposes for which this methodology was designed,

making it very suitable to achieve them. Finally, SDRM employs an iterative procedure to improve the contributions based on the experiences and information gathered throughout the research phase. As a result, this iterative procedure may assist to improve such contributions and better comprehend their true effect on the early detection of AD by the neuroradiologists.

SDRM arranges research into a five-stage methodical iterative procedure [38].The five SDRM methodological phases and their use in this dissertation are discussed below (**see Figure. 1-2**).

1. **Build a conceptual framework.** During this phase, researchers formulate a meaningful RQ, study system needs, and look for new ways and ideas in relevant areas. To that purpose, **we conducted a SLR** [39] during the first methodological cycle (**Cycle 0**), which assisted us in identifying a lack of empirical research done in clinical settings for the early prediction of AD and (ii) a lack of E2EL and 3D TL-based DL models.

   The first restriction prompted us to conduct **thirteen empirical investigations** (CONT #1) to better understand neuroradiologist demands, platform limits, and the findings of DL models utilizing neuroimaging biomarkers during Cycle 0. The results of the first cycle of research (literature review and empirical investigations) led us to define a set of traits and conditions that would make it easier to use DL-based techniques to diagnose the onset of MCI.
2. **Identification and implementation of the optimum DL model.** The next step in the SDRM process is to **offer the neuroradiologist computer-interpretable information for analyzing an MRI scan**. Eight DL models based on 3D CNN from the DenseNet and EfficientNet families have been implemented to achieve this goal. Following this implementation, I conducted a comparative study utilizing an efficiency effects graph and a rank mechanism (**Cycle 1**) to choose the optimal model. The results of this analysis revealed that the EfficientNetB0 and DenseNet264 models performed the best among all models for binary classification tasks to categorize (AD, sMCI) and (AD, CN) subjects.
3. **Assess and enhance the model .** In this step, we **proposed and implemented a novel method termed "Fusion of E2EL and TL"(Cycle 2)** to improve the accuracy of EfficienNet-B0 (identified in Cycle 1) for categorizing (AD , sMCI) subjects. This optimum model is one of the main contributions (CONT#2) of this dissertation. In

addition, for implementing the ensemble learning-based tool, we built a DL model for the multiclass classification task to categorize (AD, sMCI and CN) subjects.

4. **Deploy the model (DEEP-AD).** This phase entails the **creation of a system prototype**, DEEP-AD (CONT#3). The system prototype helped to (1) fulfill the requirements identified as essential to support the neuroradiologists; (2) show the viability of the ensemble of models (i.e., EficientNet-B0, DenseNet264); and (3) assess their usefulness in real-world scenarios (**Cycle 3**). In addition, the design of the interfaces was modified throughout the development process in accordance with the technologies employed.
5. **Observe and evaluate the DEEP-AD.** During this phase, researchers (1) **assess the system's performance and usability**, (2) confirm the system's level of support for the established criteria, and (3) monitor the effect on people. Given this context, in the first assessment iteration, a data set of **41 MRI scans of Spanish Subjects** was used, which were verified by neuroradiologists from HT Medical in Spain.

   In the second iteration, Dr. Alvero Berbis, Head of the Neurology Department at HT Medica, Spain, further validated the tool. This gave me the useful information that allows me to judge the success of the linked dissertation goals in a positive way.

1.4 Document Structure

The following is the structure of this dissertation.

Chapter 2 delves into the theoretical and practical background of the dissertation research context. This includes the main ideas behind (i) AD, including the differences between stable and progressive MCI, clinical requirements to treat AD patients on time, and current drawbacks; (ii) ML and DL approaches for early detection of AD, including the differences with other similar concepts, DL models used to automate the classification of AD, and associated orchestration tasks; and (iii) neuroimaging biomarker datasets and preprocessing methods (iv) the current state and limitations of DL strategies obtained through a systematic literature review . In addition, the dissertation's first contribution (CONT#1) is presented, which consists of thirteen empirical studies that were conducted utilizing several DL algorithms for the early detection of AD. The later sections of this chapter detail the research findings and provide possible design principles based on them.

In **Chapter 3**, we described the design and implementation of eight 3D CNN-based architectures employing E2EL and T1-weighted MRI biomarkers. The preprocessing techniques used for MRI input and the sources of MRI scans are also explained. In addition, the employed rank mechanisms, and efficiency effect graphs to compare the implemented models have also been presented. The analytical findings and empirically verified model results are outlined at the end of the chapter.

In **Chapter 4**, we enhance Chapter 3 by suggesting a unique "fusion of E2EL and TL" strategy to improve the accuracy of classifying AD and sMCI subjects. For this task, 3D CNN EfficientNet-B0 was used. In addition, we developed and implemented a 3D CNN model for multi-classification tasks for classifying AD, CN, and sMCI subjects. The chapter concludes with an overview of the analytical findings and experimentally validated model results. We have also shown the second main contribution of the dissertation (CONT#2), which is the best DL model, EfficienNet-B0, built using both E2EL and TL and having an accuracy of 93.10% for unseen data. This suggests that the model could be used in clinical settings to help neuroradiologists make decisions.

In **Chapter 5**, we introduced the **DEEP-AD**, an online tool based on ensemble learning. Ensemble is created using the models described in Chapters 3 and 4. Since it is web-based and has an intuitive interface, neuroradiologist may utilize it with ease. Screenshots and a process flowchart describing how to use this tool are provided. DEEP-AD is the subpart of (CONT#3), developed to aid neuroradiologist in taking the decision about the progression of AD.

DEEP-AD was validated in **Chapter 6** by using the 50 MRI scans of Spanish data sets given by HT Medica, Spain. The models were not trained, validated, or tested using this data set. Our precision is 90.4%. Each patient's hospitalization information, MRI procedure date and time, neuroradiologist's name, and observations are also included. We have also included the findings from each MRI scan that were verified with DEEP-AD. The validation process is also a part of the (CONT#3).

Chapter 7 summarizes the whole dissertation, focusing on the implications and significance of this work for the field of applied machine learning in health. Future research directions that have arisen from the work done for this dissertation are also highlighted in this section.

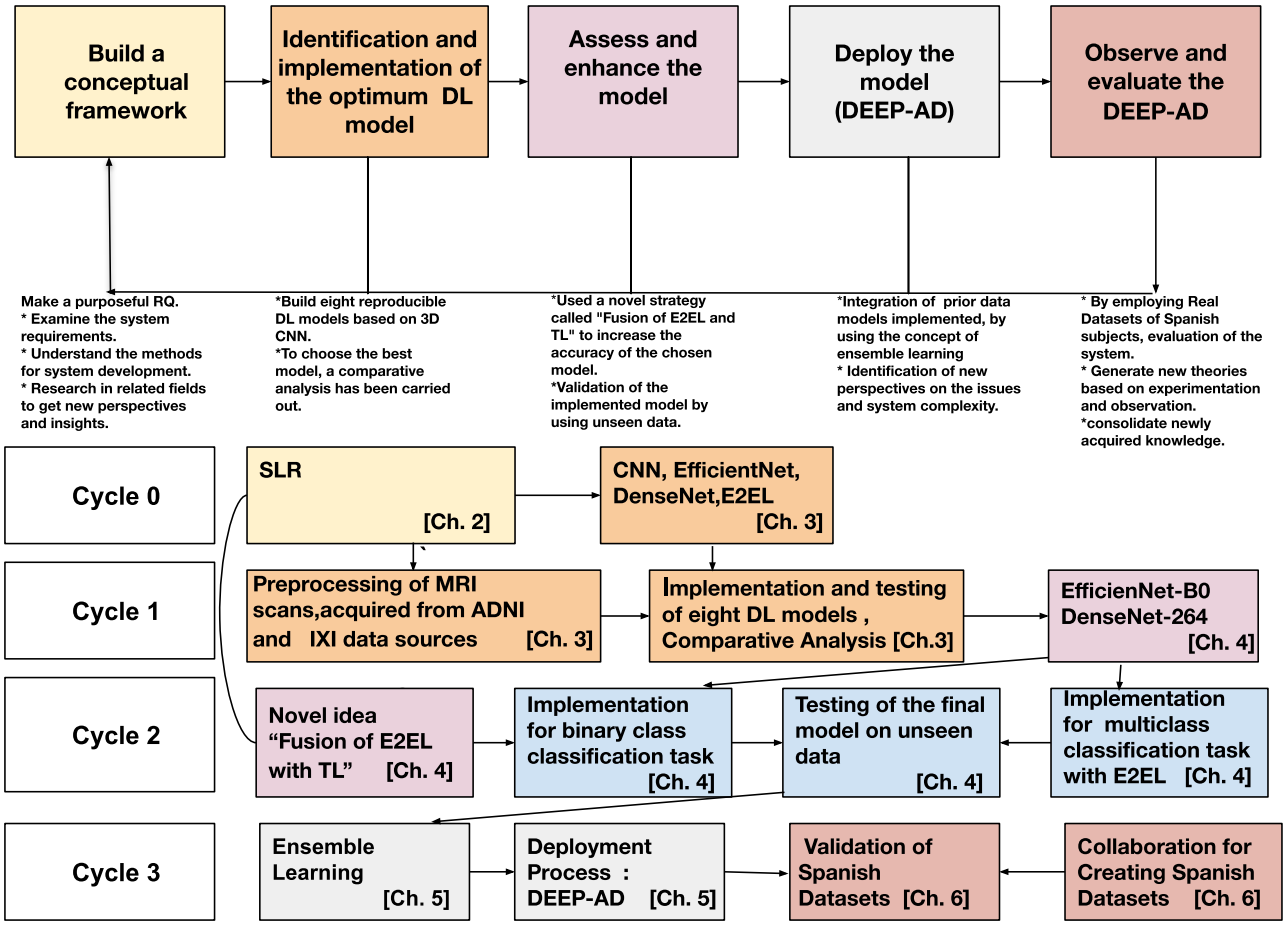


Figure 1-2 The System Design Research Methodology [38] life cycle (top), as well as the research procedure followed throughout this dissertation, including the relationship between the dissertation chapters (bottom).

Finally, the appendices contain following supplemental information.

- The pre-processing workflow and all the specifics of the DL models used in the thirteen empirical studies.
- The results of all folds of the best model for AD vs. CN, AD vs. sMCI, and AD vs. CN vs. sMCI tasks.

- Link to the codes of all the implemented models proposed in chapters 3 and 4.
- Description of the database and a link to preprocessed MRI images used during the training and validation process.
- DEEP-AD Screenshots and feedback form for neuroradiologists.
- Link to access the clinical reports and MRI scans of Spanish Datasets.
- Letter of Validation of DEEP-AD from HT Medica, Spain.

Chapter 2: Research Context

Summary: This chapter presents the dissertation's background, concentrating on the challenges mentioned in the prior chapter. First, the chapter describes the **existing landscape of early detection of AD by ML strategies, including its characteristics, advantages, and downsides**.

One of the major limitations mentioned in the literature is the absence of ICT-based tools to assist neuroradiologists in making choices concerning the course of AD, which might possibly be addressed using DL methods. As a result, the **concepts of TL and E2EL** are presented, along with their theoretical foundations and ramifications for DL models throughout the various stages of training and validation.

Then, we concentrate on **neuroimaging biomarkers and their preprocessing strategies** and available data sets, explaining the quirks that might impede the generalization of DL advantages reported in other fields. We present the results of a systematic literature review on the use of ML and DL techniques to shed light on the significance of DL tactics for early detection of ADs. Despite an increasing amount of research addressing DL in AD, the results demonstrate that there is a shortage of empirical studies done in real clinical settings, leaving insufficient evidence of the actual impact of DL-based interventions on neuroimaging biomarkers. The main problems brought up in the literature study and the solutions suggested for them are also looked at.

2.1 Introduction

The primary purpose of machine learning (ML) research in the health care domain is to facilitate and enhance decision-making regarding illness diagnosis or progression utilizing biomarkers through technology [40]–[42] . This chapter outlines the dissertation's primary pillars, which are framed within the broad domains of DL [13] and AD [43]. Both study areas have been extensively researched in the literature. However, combining these fields creates additional obstacles and research opportunities, which are also discussed in this chapter.

In the recent decade, there has been an explosion of interest in using deep learning (DL) to biological and healthcare analysis and diagnosis. Numerous successes [24], [30], [44]–[46] have been documented using the strategy for discovering relevant characteristics and completing tasks that were previously difficult to complete by other techniques and human specialists. However, specific characteristics related to early detection of AD (e.g., a massive number of dimensions and heterogeneity in neuroimaging biomarkers, similar anatomy of the brain in different stages of AD,

and the lack of empirical studies done in clinical settings) present several challenges to the development of a DL-based tool for clinical use. Based on the benefits seen in a number of health-related problems, this dissertation suggests using DL (especially techniques based on TL and CNN) for early detection of AD.

As background for this dissertation, Section 2.2 explains where AD comes from and how it differs from other health care problems that have been treated using ML methodologies. The theoretical foundations of these techniques for early AD prediction and their ramifications for clinicians are also introduced in Section 2.3, along with the notions of classic ML strategies, DL strategies using E2EL and TL, and hybrid approaches.

Nevertheless, it's important to remember that the ML distinguishing properties may affect the development of a tool for early prediction of AD (see Section 2.3). The present panorama of ML strategies for AD is not fully established; hence, we recommend carrying out a systematic literature review to collect evidence on its potential advantages (see Section 2.4). The findings highlighted two significant restrictions, both of which are recommended to be solved in this dissertation. There are two main obstacles to the effective implementation of DL-based methods in AD contexts: (1) a dearth of corroborating empirical data and (2) an absence of technology mechanisms to facilitate their orchestration in clinical settings (see **Figure. 1-1**)

2.2 Alzheimer's Disease

The cognitive and emotional declines that characterize dementia are hallmarks of the illness at the clinical level. Deterioration in memory, logic, language, and perceptual interpretation are common symptoms that have a negative impact on one's ability to go about one's everyday life. Dementia is characterized by a wide range of cognitive decline, with AD accounting for 62% of all cases. Patients with AD often struggle to remember recent events and acquire new knowledge. Memory loss, cognitive decline, and alterations in behavior worsen with AD's latter stages. Symptoms often include disorientation and wandering, distrust of loved ones, prolonged difficulty with routine chores, and difficulties with communication and motor skills [2].

At the time Alzheimer's disease is identified, the neuronal damage has already progressed enough to be irreversible [47]. Damage cannot be undone when neurons die because, unlike other cells, they do not proliferate and replace one another [2]. Dementia must thus be identified in its very earliest stages in order to slow the rate of decline [48], [49]. A patient's quality of life may be improved by receiving the appropriate care by being detected early[50].

Another thing to think about is the cost of dementia. The expenditures on health care and long-term care for those living with Alzheimer's or other dementias are significant, and dementia is one of society's most expensive diseases. Alzheimer's and associated dementias will cost the United States $321 billion in 2022, including $206 billion in Medicare and Medicaid expenditures[2]. Alzheimer's disease is expected to cost about $1 trillion by 2050 unless a medication to slow, halt, or prevent the illness is produced[2].

As a result, it is critical to create low-cost diagnostic and support systems to help minimize the rising costs associated with dementia. One of the planned approaches is to introduce e-health (the use of information and communication technology, or ICT) solutions to lower costs and make health systems and solutions more accessible to everyone [30], [51]–[53]. Prince et al. [54] highlighted the benefits of e-health technologies, such as increased access to health-care services by elderly individuals, cost-effectiveness, and efficiency in managing health resources.

2.2.3 Ways of detecting Alzheimer’s

Alzheimer's symptoms are being researched to enhance the outcomes of existing methodologies or to develop novel and more accurate diagnostic tools based on new low-cost and widely accessible technology. Methods for detecting Alzheimer's disease are divided into two groups.

2.2.3.1 Invasive

Invasive approaches need acquiring data from the patient's body through procedures such as lumbar puncture or blood retrieval. These invasive approaches aim to identify possible biomarkers that might be used to predict AD [55]. Most of them are not necessarily safe or suitable for the patient, and they may be very painful at times. On the contrary, non-invasive diagnostics are safer and more convenient throughout the diagnostic procedure.

2.2.3.2 Non-invasive

Due to the development of computer technology and the availability of new technologies, virtual environment (VE) and virtual reality (VR)-based techniques are now being examined for diagnosing AD. Immersive VE technology has significant benefits as a tool for psychological study, enhancing experimental realism and enabling impossible alterations of reality [56]. Multiple studies have demonstrated that virtual environments (VEs) can be used for neurophysiological

assessment [57], and new technologies, such as emerging head-mounted displays, multi-sensory interaction devices, and three-dimensional (3D) smart technologies, can facilitate medical tests and therapies [58]. The objective is to provide more accurate and non-invasive diagnostics at home and in nursing homes. Non-cognitive and cognitive tests may also be classified as approaches to AD detection. These treatments are non-invasive and simple to apply. Cognitive tests comprise the methods used to measure the cognitive abilities of patients.

Methods based on problem-solving exercises and questions are among the most widely used for detecting cognitive impairment. This is the standard approach used by medical professionals. The MMSE, Mini-Cog test, and Saint Louis University Mental Status (SLUMS) [59] are only some of the tools that may be used to assess cognitive function. One of the most popular and often administered assessments is the Mini-Mental State Examination (MMSE)[60] .

Orientation, registration, attention, arithmetic, memory, naming, repetition, understanding, writing, and construction are just a few of the 10 areas tested throughout the MMSE's 20 activities. While the Mini-Mental State Examination (MMSE) is sufficient for screening late-stage AD, it does not detect MCI, as Mitchell [61] argues. Since the MMSE is not difficult enough for individuals with high IQs, he also outlines some of the restrictions that are featured on the MMSE and other cognitive tests, such as the ceiling effect. When it comes to spotting signs of AD early on, the Mini-Cog test outperforms the MMSE [62].

There are a total of eleven parts to SLUMS, with a maximum score of 30 [63]. It measures things like orientation, memory, reasoning, and executive functioning and takes around seven minutes to complete. The SLUMS test has been shown to be effective in detecting Alzheimer's disease, as shown by Szcze sniak et al. [64]. The findings of this test are associated with those of other tests like the Mini-Mental State Examination. Its specificity and sensitivity for detecting MCI are also higher than those of the Mini-Mental State Examination for MCI.

Non-cognitive testing, on the other hand, comprises all other approaches used to identify and diagnose dementia. Several non-invasive, non-cognitive tests are used to find out about AD in its early stages. Some of them are discussed below.

i) **Neuroimaging Techniques :** Neuroimaging methods, such as magnetic resonance imaging (MRI) and computed tomography (CT), are used to identify disease-related brain alterations in patients [2], [65]. One of the advantages of neuroimaging is its simplicity of execution. It often requires little more than lying still and remaining

motionless while the scan is performed around you. These cutting-edge procedures allow physicians to map the brain's regions and functions noninvasively.

Agüera-Ortiz et al. [66] presents a research that uses several MRI procedures, including Diffusion Tensor Imaging (DTI) and Fluid-attenuated Inversion Recovery (FLAIR), to connect the apathy of AD patients with alterations in white and grey matter.

However, inter-individual variability in the processing of life experiences and fundamental emotions likely contributes to contradictory results across studies [61]. In neuroimaging research, it is difficult and complicated to control for symptom uniformity, nonspecific characteristics linked to psychotherapists, the subtleties of the procedures utilized, as well as the qualitative processing of subjective experiences.

ii) **Behavior Analysis:** Approaches based on the examination of behavior attempt to identify atypical responses to common situations or challenges in everyday life. Methods for assessing cognitive impairments in everyday life activities often entail the use of external equipment for a length of time in order to analyze the behavior of patients [68], [69]. For instance, Aztiria et al. [68] put a sensor on the patient's foot in order to analyze their gait (step length and step height), which represents the severity of dementia in patients. However, the gait-measuring approach has significant limitations, such as requiring the patient to wear a gadget for extended periods of time.

The primary difficulty with these approaches is that the resulting data are insufficiently precise to inform an appropriate strategy for early-stage dementia diagnosis[69], [70]. Among the disadvantages of these techniques are the need for the patients' permission to install the sensors and security issues that might lead to an invasion of privacy.

Emotional analysis is one of the additional approaches to behavior. As noted in [71], the deterioration of social cognition is one of the indications of AD; hence, several recommendations [72], [73]concentrate on the patient's capacity to recognize emotions. In contrast, some methodologies analyze patients' responses to certain inputs. Due to the significance of emotion detection in several areas, such as neuroscience and psychology, several approaches for recognizing human emotions have been presented [74]. Several approaches attempt to analyze these reactions/emotions by using various data, such as electroencephalogram (EEG), eye tracking data, audio, and facial

movements. Some dementias, such as dementia with Lewy bodies, cause a lack of facial expression; hence, facial expression analysis is implausible. However, these treatments may be especially beneficial in AD, when cognitive impairment is coupled with greater facial expression [75].

When compared to fluid biomarkers, neuroimaging techniques have a greater degree of spatial sensitivity[50], [76], [77], making them a potential tool for assessing disease development at different stages. Considering the above, we have decided to use neuroimaging methods as one of the inclusion criteria for SLR.

2.3 Machine Learning Strategies for Alzheimer's Detection

The objective of ML, which is a subfield of artificial intelligence, is to create strategies that allow computers to learn. As stated by Rogers et al.[78], it seeks to identify relationships between input factors and related responses that permit the prediction of reactions to new input variables. Using data from patients' tests and tasks or medical data and their related classifications, such as healthy or unhealthy subjects, ML approaches might establish relationships between data and labels. For instance, Souza et al. [79] uses SVM, Nave Bayes, and optimal-path forest (OPF) classifiers to distinguish between healthy and Parkinson's patient handwriting. Depending on the number of labels to be identified, cognitive impairment screening methods often use binary classification, or multiclass classification techniques. These methods are mostly used by AD detection strategies that classify vast quantities of data. The simplest classification tasks involve categorizing data into two categories, or the binary classification issue. Multiclass challenges demand classifying the data provided into more than two categories. Multiple binary classifiers or a multiclass classifier may be used to tackle this challenge.

In the last decade, it has been found that ML approaches are very useful for the diagnosis of AD [80], [81]. SVM, artificial neural networks (ANN), and DL are the approaches for classification that are most often utilized. The nature of the optimization problem is the fundamental distinction between SVM and ANN. SVM provides a globally optimal solution, while ANN provides a locally optimal solution [82]. Feature extraction is a crucial stage in both SVM and ANN. Shi et al.[83] postulated that a mix of neural networks and intelligent agents may be advantageous for medical picture processing. Nevertheless, deep learning includes feature extraction in the learning model itself. Deep learning is beneficial for huge datasets, particularly

image data [84]. Some studies have employed ensemble approaches to increase the accuracy of AD and other diseases categorization [85]–[88].

In the next section, we'll talk about how ML is used to find the early stages of AD by categorizing it in to three subtypes. This will help us understand it better and, as a result, lower our SLR inclusion criteria.

2.3.1 Traditional Machine Learning Methods

SVM is a very robust and commonly used ML approach for regression and classification problems [7], [89]. SVM employs the notion of maximum margin to categorize the data points. Many versions of SVM have been developed to improve its generalization ability and training duration. Twin support vector machine (TWSVM) and least-squares-based (LSTSVM) algorithms [90] are computationally efficient variations of SVM. Traditional ML algorithms place a significant emphasis on feature selection when classifying data.

Vemuri et al. [91] discovered that combining demographic and genetic data with MRI scans improved the accuracy of distinguishing between CN and AD. A new parcellation approach for identifying small gray matter alterations is presented by Mesrob et al. [92]. Magnin et al. [93] introduced a technique for selecting features based on the histogram of areas of interest (ROIs) for CN vs. AD. Gerardin et al. [94] used shape characteristics of the hippocampus to differentiate between CN, MCI, and AD and discovered that shape deformation features are superior to volumetric features. Other approaches, such as SVM-RFE [95], are employed to identify significant brain characteristics for CN vs AD. EEG information is also used[96] for classifying CN vs AD using SVM.

By giving researchers from all across the globe a common platform, competitions like the Alzheimer's disease Big Data Dream Challenge[2] and the challenge of predicting MCI from MRI data [3] have shown their utility in the AD diagnosis process. Numerous ML algorithms have been developed because of these contests and evaluated. It is necessary to have a preset architectural design to employ these ML algorithms for classifications. A classification method must be implemented once the features have been extracted, chosen, and the dimensions have been

[2] https://phidatalab.org/news/alzheimers-disease-big-data-dream-challenge/

[3] https://www.kaggle.com/c/mci-prediction

reduced, which amounts to four stages in total, which may be time-consuming. Additionally, each step of this process must be optimized, and professionals must be involved [13]. These implementations' reproducibility has been a problem [14].

2.3.2 Deep Learning

Since 2014, there has been a considerable amount of interest in research on AD detection using DL. Figure 2-1 shows that the number of publications in this field has grown dramatically since 2017. It has been found that deep models are better at finding AD than traditional ML methods.

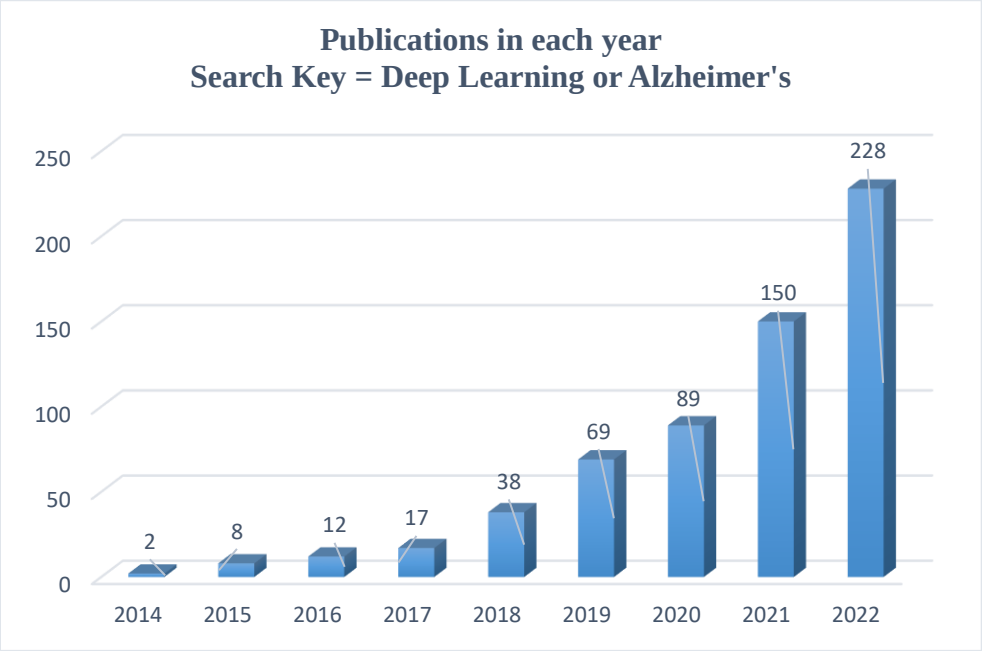


Figure 2-1 Number of papers used deep learning to find AD over the years.

Litjens et al.[97] did a review of how DL methods can be used to analyze medical images. Even though deep learning models are called "black boxes," it is said that some statistical techniques can be used to estimate the network's uncertainty. Shen et al. [84] did a survey on how deep learning could help people with Alzheimer's.

The issue with conventional ML approaches has been resolved by DL (7), which has transformed the process of feature engineering into a learning step. In other words, instead of manually extracting features, DL merely requires a data set and, if required, minimal preprocessing. It then discovers on its own the useful representations [13] . Thus, feature engineering is now performed by computers rather than humans. This indicates that non-experts

in machine learning may use DL for their own research and/or applications, particularly in medical picture analysis. Deep learning's exceptional success is largely attributable to the following factors:

- Technological advancements in central processing units and graphics processing units.
- The availability of vast quantities of data.
- innovations in DL algorithms [30], [41], [50], [98]

DL is technically superior to ordinary artificial neural networks because it permits the building of networks with numerous (more than two) layers. Hierarchical feature representations that allow higher-level characteristics to be generated from lower-level features may be discovered using deep neural networks [13]. DL has achieved record-breaking performance in a range of artificial intelligence applications and grand challenges [99] due to the ability of these algorithms to build hierarchical feature representations exclusively from data. Specifically, advancements in computer vision spurred the application of DL in medical image analysis, including image segmentation, image registration, image fusion, image annotation , computer-aided diagnosis, and prognosis[13].

When a large number of examples are provided during the training phase, DL techniques are particularly successful. In the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), for instance, there were more than one million labeled images offered[100] . However, in most medical applications, there are far fewer images. Therefore, the restricted amount of training examples available to develop deep models without overfitting is a significant obstacle in the application of DL to medical images.

Researchers have developed a variety of strategies to address this issue, such as using either 2D or 3D image patches rather than the full-sized images as input to reduce input dimensionality and, consequently, the number of model parameters; expanding the data set by artificially generating samples via affine transformation (i.e., data augmentation) and then using the expanded data set to train their network from scratch (end to end learning) [101]; Instantiation model parameters using those of pretrained models from nonmedical or natural images(transfer learning)[24], then fine-tuning network parameters using task-relevant data.

In terms of input formats, deep models may be categorized as multilayer neural networks that accept vector-format (i.e., unstructured) data and convolutional networks that accept 2D or 3D (i.e., structured) values. CNNs have got a lot of attention in the field of medical image analysis

because of how images are structured [98]. The structural or configurational information in nearby pixels or voxels is also an important source of information.

Overfitting has also played a major role in DL history [102], with attempts being made to tackle it at the architecture level. One of the earliest models devised to tackle the overfitting issue was the Restricted Boltzmann Machine (RBM). Stacking RBMs led to the formation of deeper structures known as Deep Boltzmann Machines (DBM). The Deep Belief Network (DBN) is a supervised learning system that uses data extracted from each stacked layer to link unsupervised characteristics [103]. DBN was discovered to outperform other models, which is one of the reasons why deep learning has grown in popularity. While DBN tackles the overfitting issue by employing RBM to lower weight initialization, CNN efficiently reduces the number of model parameters by introducing convolution and pooling layers, resulting in a reduction in complexity. CNN is frequently employed in the area of visual recognition because of its efficacy when given adequate data. Transfer learning (TL), which allows DL training to be successful in the situation of limited data, has emerged as being quite popular in the area of DL, according to recent research [26] . TL is suggested when compared to human behavior since we may use the information, we've acquired to solve new, challenging situations.

When compared to distinguishing between progressing MCI and stable MCI, the classification of AD and normal controls is straightforward. Furthermore, compared to the first task, the quantity of neuroimaging data for the categorization of stable MCI vs. progressive MCI (sMCI vs. pMCI) is much smaller, yet both task types shared similar types of biomarkers [50].

Numerous studies have used local TL because of this impression. The purpose of local TL is to classify sMCI vs. pMCI using the final weights from the AD vs. CN classifier as the initial weights. Additionally, several studies employed pre-trained 2D CNN-based architectures like VGG16 and ResNet [24], [104], [105] to initialize the weights for the classification tasks and achieved higher results. TL aids in accelerating the training process and enhancing the functionality of DL architectures. When deep learning is used together with traditional machine learning methods, i.e., SVM as a classifier, it is referred to as a "hybrid method."

The following points outlines the problems with the traditional ML, DL, and hybrid approaches for utilizing neuroimaging data to diagnose the early stages of AD. **Table 2-1** summarizes some of the highest reported accuracy by ML, DL, and hybrid approaches [41], [50]. While standard ML techniques like SVM have shown promising results, DL techniques like CNN

and sparse autoencoders have outperformed ML techniques[50]. In hybrid approaches, classic ML techniques like SVM is used with DL for feature extraction and classification[106]. For the early identification of AD (sMCI vs. pMCI task), the SVM for classification and the stacked autoencoder (SAE) for feature selection obtained accuracy levels of up to 83.3%. Maximum reported accuracy for the DL approach is 82.4%, while for the ML method is 82.5% [41], [50], [106]. Traditional ML and DL models have a limitation in that they only use samples from a particular domain, which affects their accuracy when the number of samples is small [13]. TL makes use of samples from both the target domain and several auxiliary domains[106].

Methods	**Algorithm**	**AD vs. CN**	**sMCI vs. pMCI**	**Ref.**
Machine Learning	Linear-SVM	88.90%	70.70%	[7]
	Temporally Structured (TS)-SVM	-	82.50%	[8]
	Linear-SVM	93.01%	75.00%	[9]
	Least Absolute Shrinkage and Selection Operator (LASSO)- SVM	95.10%	65.40%	[10]
	Radial Basis Function (RBF) - SVM	81.25%	69.23%	[11]
Deep Learning	CNN	92.87%	76.21%	[107]
	Deep Boltzmann Machine (DBM)	90.00%	78.00%	[88]
	Deep Neural Network (DNN)	84.60%	82.40%	[108]
	CNN	91.09%	76.90%	[109]
Hybrid	SAE & SVM	98.80%	83.30%	[110]
	DBM & SVM	93.52%	74.58%	[111]
	RBM & SVM	91.40%	57.40%	[112]

Table 2-1 Accuracies reported for binary classification tasks for classifying different stages of AD

The hybrid methods produce good results with a limited amount of data, but they do not fully utilize DL [41]. The concept of TL employing pre-trained CNN networks entered the picture at this point. Pre-trained CNN networks are trained on datasets including millions of images over the course of several weeks on multiple servers with the assistance of experienced specialists. By fine-tuning the last layers of the CNN, these networks can be utilized with tiny datasets[98].

Considering all of this, we chose DL algorithms implemented with TL as one of the most important criteria for SLR.

2.4 Systematic Literature Review

I have conducted a SLR that included papers published in the previous ten years. The studies covered in this review concentrate on DL architectures, TL, and neuroimaging data to classify the different stages of AD. SLR was carried out in stages in accordance with **PRISMA [113] methodology**.

The primary goal was to discover gaps in the state-of-the-art ML applications in the AD research domain. The research questions (RQs) were first identified. Second, the search strategy was implemented, and the inclusion and exclusion criteria for picking relevant articles were defined. Finally, data extraction was performed to respond to the RQs. Furthermore, responses to these RQs were provided while emphasizing the field's challenges, limitations, and future potential. These steps were discussed in the sub-sections that follow.

2.4.1 Research Questions

Examining the critical components of the current literature is required to support future research activities. The goal is to contribute to the development of a better decision-support system for the early detection of AD. Following an anticipatory data reduction approach, the SLR was framed into eight separate research questions (RQs) to steer this process [114].

- RQ1: Which DL models and TL strategies have been utilized to capture AD-related patterns?
- RQ2: Which neuroimaging biomarkers and parameters are utilized?
- RQ3: Which pre-processing methods are used to deal with neuroimaging biomarkers, and how is input for DL models managed?
- RQ4: What are the current levels of accuracy and other performance indicators?
- RQ5: What are the publicly available data sources?
- RQ6: Which software systems are utilized to pre-process neuroimaging data and execute deep learning algorithms?
- RQ7: What methods have been used to lessen overfitting?
- RQ8: What are the ongoing research opportunities, gaps, and challenges?

2.4.2 Methodology

Search Procedure: The Science Direct, IEEE Xplore, Web of Science, ACM Digital Library, and PubMed databases were considered. These databases were chosen because they are

the most reliable sources and include the pertinent peer-reviewed papers. The following search term was used: “Deep Learning” & “Alzheimer” & “Neuroimaging” & “Transfer Learning”.

Sysrev [115] was used for the search. This tool enables collective data abstraction from online-accessible documents and supports it by creating article filters using machine learning. **Figure 2-2** displays the relevance of significant phrases found by Sysrev for our search; early and late MCI, which is luckily also of utmost importance, is given the most weight. There were 215 total articles obtained, including 161 from Science Direct, 15 from the Web of Science, 6 from PubMed, 25 from ACM, and 8 from IEEE Xplore. The inclusion and exclusion criteria were used to further evaluate and analyze the papers. Redundancy was carefully avoided throughout the filtering process. Sysrev aided with the screening process.

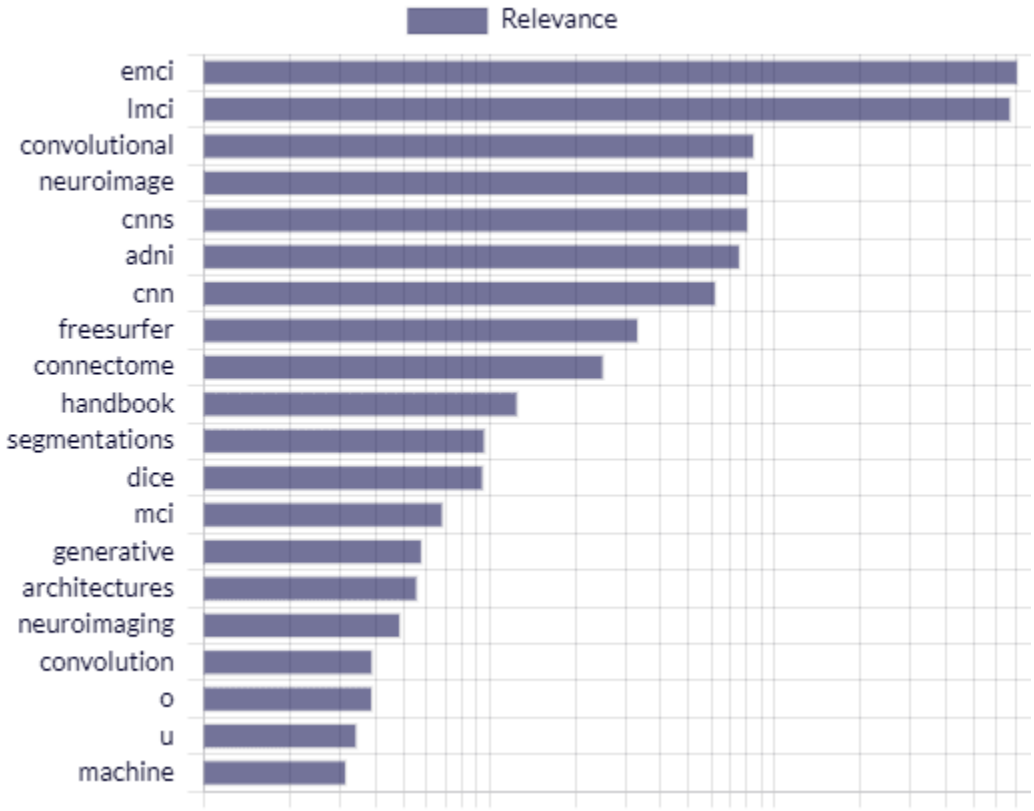


Figure 2-2 Significance of key phrases determined through Sysrev

Criteria for Inclusion and Exclusion: The selection of publications that are important and consistent with the research goals depends heavily on the inclusion and exclusion criteria. For the diagnosis and early detection of AD, we considered articles that used DL architecture with TL and neuroimaging biomarkers. Cross validation and at least two prediction matrices were also

considered. One of the criteria for inclusion was the thorough description of the data sources and pre-processing techniques used for the biomarkers. Survey papers, dissertations, patents, handbooks, and encyclopedia entries were not included.

Study Assortment: The PRISMA flow diagram, shown in **Figure 2–3**, was used to outline the search strategy and research selection. 215 articles were found after the articles were located using the search approach. In total, 15 duplicates were eliminated. The remaining 200 items underwent a second screening. The titles and abstracts of the research were used to determine their significance. 165 publications were excluded because they did not satisfy the inclusion criteria as shown in the PRISMA flow diagram. Out of the 35 articles, 12 [116]–[127] were excluded because either they did not perform the classification task sMCI vs. pMCI / sMCI vs. AD or did not employ TL techniques. Due to the lack of at least two prediction matrices or not implementing cross-validation during the training, five articles [128]–[132] were excluded. Two articles [133], [134] were excluded because they didn't provide the necessary insights for input management.

The remaining two articles [133], [135] were excluded because they did not use neuroimaging data. Lastly, the research provided by [136] was excluded since it attempted to differentiate between AD and Parkinson's. The remaining thirteen articles [19], [22], [137]–[147] were included in this systematic evaluation. These studies are pertinent to addressing the questions posed at the beginning of this section. The Journal of Neuroscience Methods, Neuroimage, Clinical and Behavioral Brain Research, and Medical Image Analysis are the top journals in this field of study. According to Google Scholar, the most referenced articles are [137], [144], [145], [147].

Nine studies from Science Direct, two from IEEE Xplore, one from PubMed, and one from Web of Science were included. We found only eligible published articles between 2017 and 2020 according to inclusion criteria—one study in 2017, five studies in 2018, three studies in 2019, and four studies in 2020. **Table 2-2** contains information on the included work, year, journal name, database, country, and citations.

Risk of bias: The primary drawback of undertaking a systematic evaluation of the literature is that it may be biased. To prevent this situation as much as possible, a search key has been created by combining keywords and logic operators. In accordance with the PRISMA principles, I have attempted to adhere to the most optimal review criteria and procedures. Early researchers published a number of reviews[41], [50], [106] of ML and DL approaches for the categorization of AD.

However, they did not cover all these RQs; TL methodology-related procedures were notably absent.

Jo et al. [41] did an in-depth study of research published between 2013 and 2018, comparing the diagnostic classification accuracy of pure DL and a mix of DL and ML approaches. They also looked at the variation in accuracy based on the type of biomarker. Tanveer et al. provided a review of papers published between 2005 and 2019 and performed an in-depth analysis based on three categories, namely state vector machine, artificial neural network, and deep learning models. Ebrahimighahnavieh et al. provide the most recent findings and advancements in the field of AD detection using DL. Specifically, valuable biomarkers and characteristics, pre-processing procedures, and various approaches to dealing with neuroimaging data from single-modality and multi-modality experiments.

The primary distinction between previous studies and this SLR is that they did not do a comprehensive analysis of the change in accuracy based on the DL and TL combinations, as well as the kind of biomarker and input management approach. As a result, this work may assist in determining the most successful combination of the DL model with TL and in selecting the most appropriate preprocessing approach. Furthermore, in the previous studies, there was no consideration of new biomarkers such as amyloid-PET and tau-PET [65]. This SLR study investigated the use of these biomarkers in future research. The answers to all the RQs are covered in detail in the next section. **Appendix A** contains technical insights, datasets, and a category-wise study of several pre-processing strategies for neuroimaging biomarkers, as well as the software tool utilized for pre-processing.

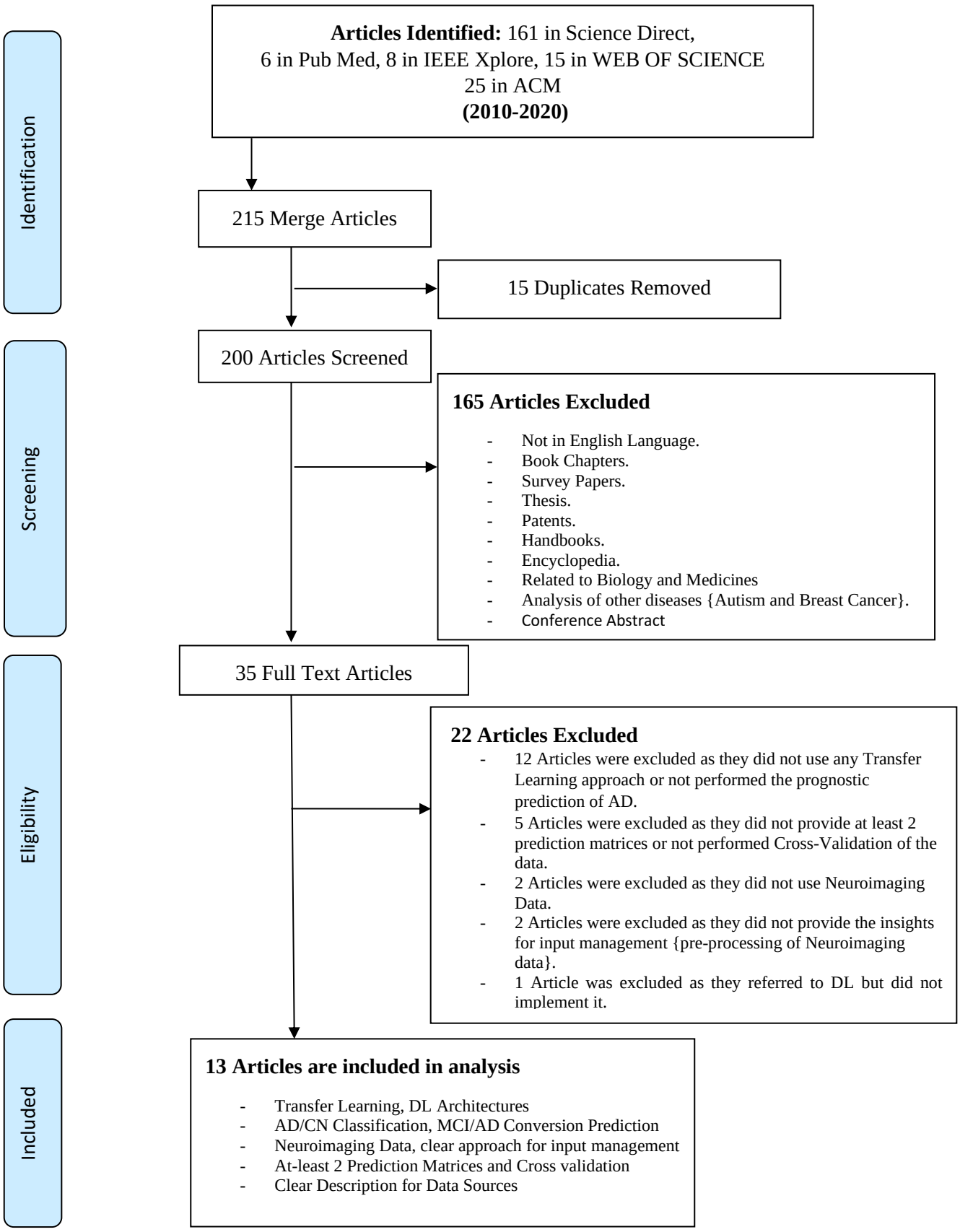


Figure 2-3 PRISMA flow diagram

Ref.	Year	Journal Name	Database	Country	Citations
[147]	2017	International Conference on Bioinformatics and Biomedicine (BIBM)	IEEE Xplore	Canada	9
[140]	2018	ELSEVIER / NeuroImage: Clinical	Science Direct	Saudi Arabia	4
[19]	2018	NCBI/Quantitative imaging in medicine and surgery	PubMed	China	3
[143]	2018	DEStech Transactions/ Computer Science & Engineering	IEEE Xplore	China	1
[144]	2018	ELSEVIER / Medical Image Analysis	Science Direct	Canada	39
[145]	2018	ELSEVIER / Behavioural Brain Research	Science Direct	Korea	55
[137]	2019	Springer/Journal of Medical Systems	Science Direct	Italy	47
[141]	2019	natureresearch/ scientific reports	Web of Science	Korea	12
[138]	2019	ELSEVIER / Journal of Neuroscience Methods	Science Direct	China	10
	2020	ELSEVIER / Journal of Neuroscience Methods	Science Direct	USA	1
[46]	2020	ELSEVIER / Medical Image Analysis	Science Direct	France	1
[52]	2020	ELSEVIER /Saudi Journal of Biological Sciences	Science Direct	China	1
[55]	2020	ELSEVIER / NeuroImage: Clinical	Science Direct	USA	0

Table 2-2 Publication details of the included articles

2.4.3 RQ1. Deep learning and Transfer learning architectures

The first RQ is about the kinds of DL architectures and TL techniques that have been implemented for the early detection of AD by using neuroimaging biomarkers. To find the answer to this RQ, the author has analyzed the DL algorithms based on the types of learning employed, as discussed below.

1. **Unsupervised DL:** To deal with data scarcity and high dimensionality, unsupervised learning can be used [102]. It is utilized to achieve a task-specific representation, which is especially useful when there is limited labeled data but a large quantity of unlabeled data. In such situations, autoencoders (AEs) are the most often employed technique. AE variants such as CAE[4], ICAE[5], and SAE[6] [26], [84] have been employed for feature extraction, dimension reduction, and **TL**.
2. **Supervised DL:** Supervised DL is comparable to the human learning concept. Well-labeled data are used to train supervised DL architectures. It teaches the learning algorithm how to apply what it has learned from the training data to new scenarios[13], [28], [102]. The model is evaluated on a portion of the testing set to predict the output when the training procedure is complete. As a result, datasets with accurate inputs and outputs are essential since they speed up the model's learning.

   Overfitting has played a significant role in the development of DL [102] , and attempts have been made to resolve it at the architectural level. The RBM[7] was one of the first models created to address the overfitting issue [148]. The DBM[8] was created by stacking RBMs, which led to the formation of deeper structures [149] . Deep Belief Network (DBN) is a supervised learning technique for connecting unsupervised features by pulling data from each stacked layer (Hinton et al., 2006). DBN was discovered to outperform other models, which is one of the reasons why deep learning has become so popular (Bengio, 2009). CNN efficiently decreases the amount of model parameters by introducing convolution and pooling layers, which leads to a reduction in complexity, while DBN tackles the overfitting issue by decreasing the weight initialization using RBM. CNN is

[4] Convolutional autoencoder
[5] Inception CAE
[6] Stacked AE
[7] Restricted Boltzmann Machine
[8] Deep Boltzmann Machine

frequently employed in visual recognition as a result of its efficacy when given sufficient data. DNNs[9] have exhibited amazing performance in challenging machine learning tasks including image classification and voice recognition[150]. CNNs have a much less number of parameters than DNNs due to shared weights and pooling layers[50], [102], [142].Another model employed is RNNs [41], it can describe the temporal behavior of a time series effectively, with the output dependent on the previous calculations. RNN has been used with CNN in some studies. The need for a big dataset might be seen as a shortcoming of these models [76].

3. **Transfer Learning:** Conventional ML/DL models employ just data from a single domain, which has a significant impact on their performance when the number of accessible samples is small. Transfer learning is a method that incorporates not just examples from the target domain, but also examples from a few auxiliary (similar) domains.

Oh et al. [141] used ICAE to extract sparse representations from MRI scans and use the encoded visual features for a binary classification task of AD vs. CN by using supervised learning, and then transferred the learned knowledge to classify (sMCI and pMCI tasks). They used unsupervised learning, supervised learning, and TL together. Donghuan Lu et al.[144] employed SAE to pre-train each of the proposed multiscale DNNs. They trained a single hidden layer at a time using greedy layer-wise training [76]. This was followed by supervised fine-tuning to develop the classifier to discriminate between pMCI and sMCI.

Although many researchers opted to create their own CNN structures, some of them also make use of well-known pre-trained 2D CNN structures with TL, such as ResNet, deep-ResNet, CaffeNet, AlexNet, DenseNet, VGG16, GoogleNet, and Inceptionv4. ResNet-18 is used by Ramzan et al. [137] to initialize the weights of their 2D CNN before training the network. CaffeNet is used by Lin et al. [143] and Yang et al. [146] used the AlexNet. In multiple classifications and estimates of conversion risk, Wu et al.[19] has implemented two CNN architectures, GoogleNet and CaffeNet. with TL. A customized version of ResNet was suggested by Abrol et al. [22] to evaluate neuroimaging data and predict the progression of MCI to AD. Two distinct pre-trained architectures, VGG16 and Inception V4, were employed by Hon et al [147]. An age-adjusted 3D CNN-based pre-training model was proposed by Gao et al[139] . Four distinct CNN

[9] Deep Neural Network

architectures—3D subject-level CNN, 3D ROI-based CNN, 3D patch-level CNN, and 2D slice-level CNN—were proposed by Junhao Wen et al.[142]. Also, two different methods were used for transfer learning: ResNet for 2D CNNs and AE pre-training for 3D CNNs. Silvia Basaia et al. [140] used a 3D CNN to classify AD vs. CN using E2EL (training from scratch). They then used the final weights of the AD vs. CN classifier as the initial weights of the sMCI vs. pMCI classifier. Li et al. [138] presented a hybrid method for analyzing the hippocampus that used both convolutional and recurrent neural networks. The main points are summed up below.

- For extracting features from neuroimaging biomarkers, the unsupervised DL architectures SAE, CAE, and ICAE can be utilized. They may be used to determine the starting weights for CNNs. A shortcoming of AEs is that they learn to record as much information as they can, not just pertinent information. Pretrained 2D CNN models can be also used to integrate the weights as well, but 3D data must first be converted into 2D slices to utilize them.

- Most of the studies use CNN based supervised algorithms, which make it possible to combine feature selection and classification into a single algorithm. 3D CNNs demonstrated superior performance in the extraction of local features. However, the complexity of training is a restriction that may be overcome using ROI or patch-based pre-processing techniques. 2D CNN, on the other hand, is simpler to train but insufficient for storing the spatial information of 3D data.

2.4.4 RQ2. Neuroimaging Biomarkers and other parameters

The second RQ addresses the neuroimaging biomarkers and the characteristics that are used to categorize and forecast the development of AD. Non-invasive biomarkers have been widely employed, including MRI and PET[19], [22], [132]–[142]. MRI makes it possible to examine the abnormal brain changes associated with AD, it creates a 3D representation of the bones and soft tissues using a magnetic field and radiofrequency pulses [151].

There have been several MRI scan types employed. In [19], [22], [143], [147], structural MRI (sMRI) has been used. The capacity to trace brain shrinkage and aid in identifying a potential AD etiology are two additional benefits of sMRI, in addition to its excellent spatial resolution and accessibility [66], [152]. Blood flow variations are also shown by functional MRI (fMRI). Resting-state fMRI (rs-fMRI) records changes in patients' blood oxygenation levels when they are at rest. As a consequence, different patterns of blood oxygenation levels may be seen in the areas of the

brain affected by neuronal degeneration [12], [18], [153]. The fMRI was used in one of the included studies [137].

Several authors have used T1-weighted scans (T1W). T1W scans, which are a component of MRI protocols, are considered the most anatomical scans[10]. One of the most important indicators for identifying AD progress are GM, WM, and CSF[11] measures [154]. T1W scans provide the closest representation of these measures[12]. In [138]–[142] T1W scans has been used. Different forms of T1W scans "3T/1.5T" refers to a 3Tesla/1.5Tesla MRI [13], which is created by a magnetic field of either 3Tesla or 1.5 Tesla and produces a clearer and more comprehensive picture. Most T1W scans are captured by the magnetization-prepared rapid gradient-echo (MP-RAGE) [155] procedure to capture strong tissue contrast and high spatial resolution with whole-brain coverage.

A functional imaging technique in nuclear medicine called positron emission tomography (PET) [65], [156] is used to monitor the metabolic process in order to find AD. The most common form of PET scan is fluorine-18 fluorodeoxyglucose (18F-FDG) PET [157]. It gives a measurement of how well cells process glucose. It can help with the neurocognitive problems brought on by AD [77]. Since it may show the characteristic patterns of AD sooner than MRI for MCI people, it is primarily useful for the early identification of AD [65]. The assessment of brain amyloid deposition, one of the key neuropathological milestones of AD, is done using amyloid-PET, a different form of PET [65], [157].

To increase the precision of AD diagnosis, several biomarker combinations were applied in many studies. Lu et al. [144] combined 18F-FDG-PET with sMRI, while choi et al. [145][66] combined 18F-FDG-PET with AV-45 (florbetapir) PET. Yang et al. [146] exclusively use 18F-FDG-PET.

Tau-PET imaging is an additional PET type [158]. Abnormal tau accumulation is a major contributor to neurodegenerative diseases such as Alzheimer's [65]. The development of tau-specific PET tracers like THK5351 and THK5357 has been guided by improvements in neuroimaging technology [159]. We did not find any articles where authors used tau-PET to predict the course of AD using DL and TL techniques.

[10] https://radiopaedia.org/articles/mri-sequences-overview
[11] GM: Gray Matter, WM : White Matter, CSF: Cerebrospinal Fluid
[12] https://case.edu/med/neurology/NR/MRI%20Basics.htm
[13] https://www.mana.md/what-is-3t-mri/

Gender, age, speech pattern, EEG, tau protein, Aβ protein, retinal abnormalities, postural kinematics analysis, MMSE and CDR score, logical memory test, and several genes thought to be important for AD are a few more characteristics that are relevant to AD identification [6]. Mean subcortical volumes, gray matter densities, cortical thickness, brain glucose metabolism, and cerebral amyloid-β accumulation are the most often selected characteristics from these biomarkers during ROI-based analysis.

Figure 2-4 depicts healthy and AD biomarkers. Left to right, their clinical acceptability decreases. In the sMRI, grey matter volume is displayed in blue, in the 18F-FDG PET images, decreased metabolism is displayed in green, in the amyloid-PET images, the small amount of amyloid deposition is displayed in green or blue, and in the tau-PET images of AD patients, low tau-tracer retention is displayed in green or blue[65], [98]. TSPO[14]-PET [159] and SV2A[15]-PET [160] are two more types of potential PET scans that can be employed in future research. TSPO-PET scans are shown as yellow or red axial slices, indicating neuroinflammation due to enhanced TSPO expression. The SV2A-PET scans show yellow or red, indicating normal synaptic density, and green, indicating decreased synaptic density.

Figure 2-5 summarize the distribution and maximum accuracy by the type of neuroimaging biomarkers, respectively. 70% of studies have used sMRI [19], [22], [138]–[143], [147], maximum accuracy achieved by combining 18-F FDG with AV-45 PET[145].

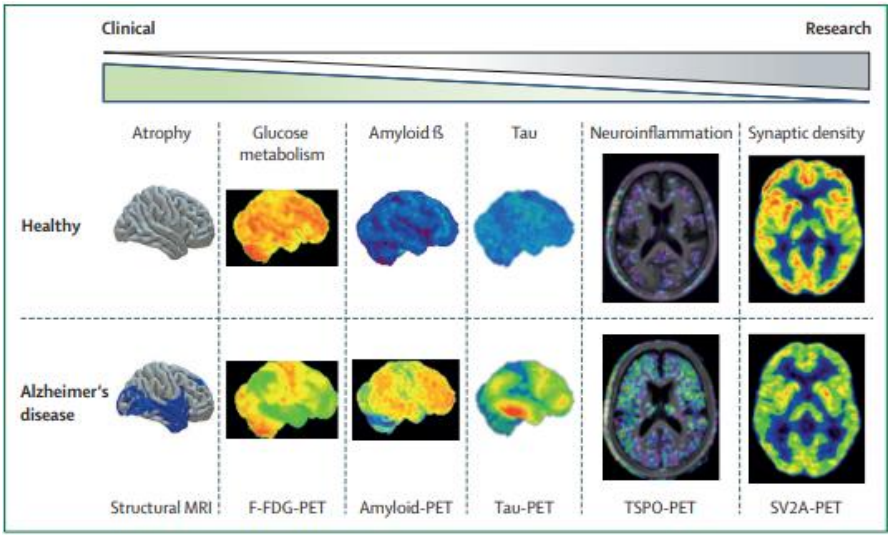


Figure 2-4 MRI and PET by distinct radiotracers [65] [98]

[14] translocator protein
[15] Synaptic vesicle glycoprotein 2A

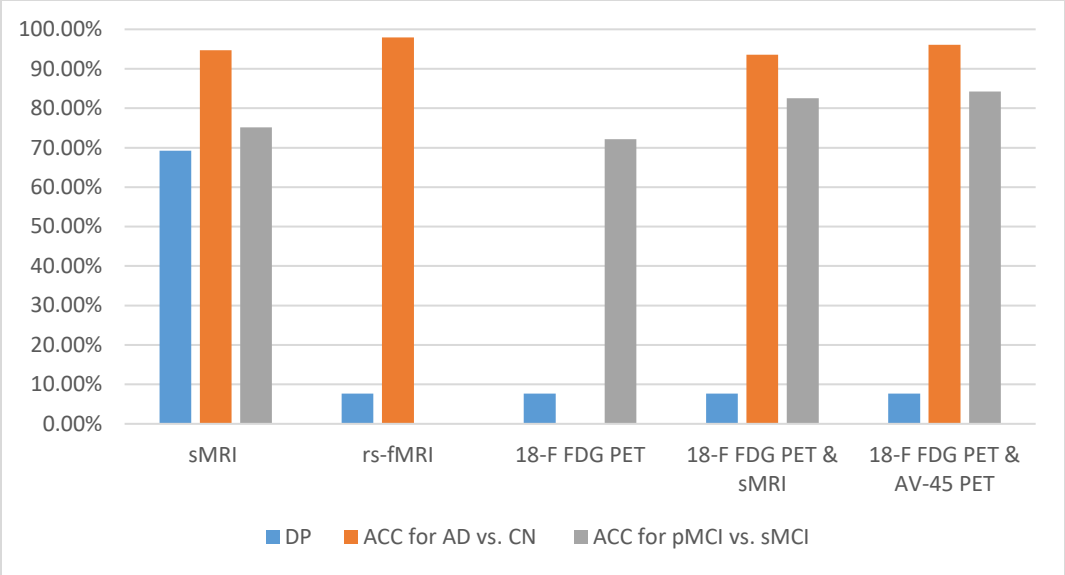


Figure 2-5 DP: Distribution Percentage, ACC: Accuracy by the type of biomarkers

2.4.5 RQ3. Pre-processing of Neuroimaging Biomarkers

The third RQ is concerned with the preprocessing approaches and input management for neuroimaging biomarkers. Preprocessing is one of the most influential determinants of a classifier's performance. It eliminates artifacts and noise from the data to increase the quality of the picture and boost feature extraction [50]. The included studies used some of the following pre-processing methods:

Normalization of Intensity: The employment of various scanners and settings to examine people at different periods might result in significant changes in intensity. It may degrade future processing performance, including registration, segmentation, and tissue volume measurement. Normalization entails mapping the intensities of all voxels against a standard scale such that similar structures have equivalent intensities [50]. Utilizing the non-parametric, non-uniform intensity normalization (N3) algorithm is the most common strategy. This is used to address the non-uniformity of the intensity by sharpening the histogram peaks [161]. It has been implemented in [19], [137], [138]. Oh et al. [141] standardized the MRI scan intensity to the interval [0, 1]. The brainstem area was employed for internal normalization in[145] since it is not predicted to be impacted by AD; thus, the mean intensity in the brainstem region is determined and used to disperse the average intensities.

Spatial Smoothing: A Gaussian filter between 5- and 8-mm lowers noise and retains visual signals [16]. In three of the listed investigations[22], [137], [143], it has been used.

Registration (also called "spatial normalization") is the process of mapping the neuroimaging scan to a reference brain space. It enables comparisons between people with diverse anatomies [50]. It standardizes neuroimaging scans in accordance with a standard template, such as MNI [17]. In [137], a linear transformation with 12 DOF (including translation, scaling, shear, and rotation) was implemented. Several articles [140], [141], [143] have used the DARTEL[18] registration [162]. Using the SyN[19] method [163], linear (affine) registration was carried out in [22]. In [142], the authors have used non-linear registration by using the unified segmentation strategy [164]. In [144], the authors registered images using a non-rigid [107] registration approach LDDMM[20].

Co-registering various modalities is an additional registration method described in the literature. There are two major landmarks in the brain[21]: the anterior commissure (AC) and the posterior commissure (PC). The AC-PC line has been adopted as a standard by the neuroimaging community and is often used as the reference plane for axial imaging to compare participants [165]. Choi et al. [145] registered PET scans such that the anterior and posterior axes of the patients were perpendicular to the AC–PC line. Li et al.[138] used the affine registration approach to align MRI scans linearly to a template. Gao et al. [139] executed strict registration in the MNI-152 space to assure orientation and position consistency. The authors of [141], [143] used interpolation techniques[22] to convert data into voxel sizes and dimensions that were identical.

Gradwrap: Restore geometry that has been distorted by gradient nonlinearity [166]. One research has employed it [19].

Tissue segmentation: Divide a neuroimaging modality into segments corresponding to various tissues such as GM, WM, and CSF in the case of MRI and probability maps and metabolic intensities of the ROI in the case of PET scans. This procedure has been used in [22], [140], [143], [144], [146] articles.

[16] Spatial Normalization - an overview | ScienceDirect Topics
[17] About the MNI space(s) – Lead-DBS
[18] Diffeomorphic Anatomical Registration Exponentiated Lie Algebra
[19] symmetric image normalization method
[20] Large deformation diffeomorphic metric mapping
[21] https://radiopaedia.org/articles/anterior-commissure-posterior-commissure-line-1

[22] Interpolation Methods For Image Registration (Biomedical Image Analysis) (what-when-how.com)

Bias field correction: The bias field signal, which is a low-frequency and very smooth signal, taints MRI images, particularly those made by outdated MRI equipment [167]. Results from DL methods that employ the grayscale values of picture pixels will not be adequate. Before presenting distorted MRI images to such algorithms, a pre-processing step is required to adjust for the bias field signal. The N41TK [168] algorithm was used in three articles [19], [141], [142] to carry out this operation.

Brain extraction (Skull Stripping): The removal of tissues other than the brain from the neck and skull. Four articles [137], [138], [142] have used it.

Motion Correction: Used to reduce the effect of head movements and improve their precision during data collection. One article [137] made use of it.

High-pass filtering: It is used to get rid of low-frequency noise signals that are produced by psychological artifacts. One article [137] made use of it.

The percentage of preprocessing techniques used is depicted in **Figure 2-6**.

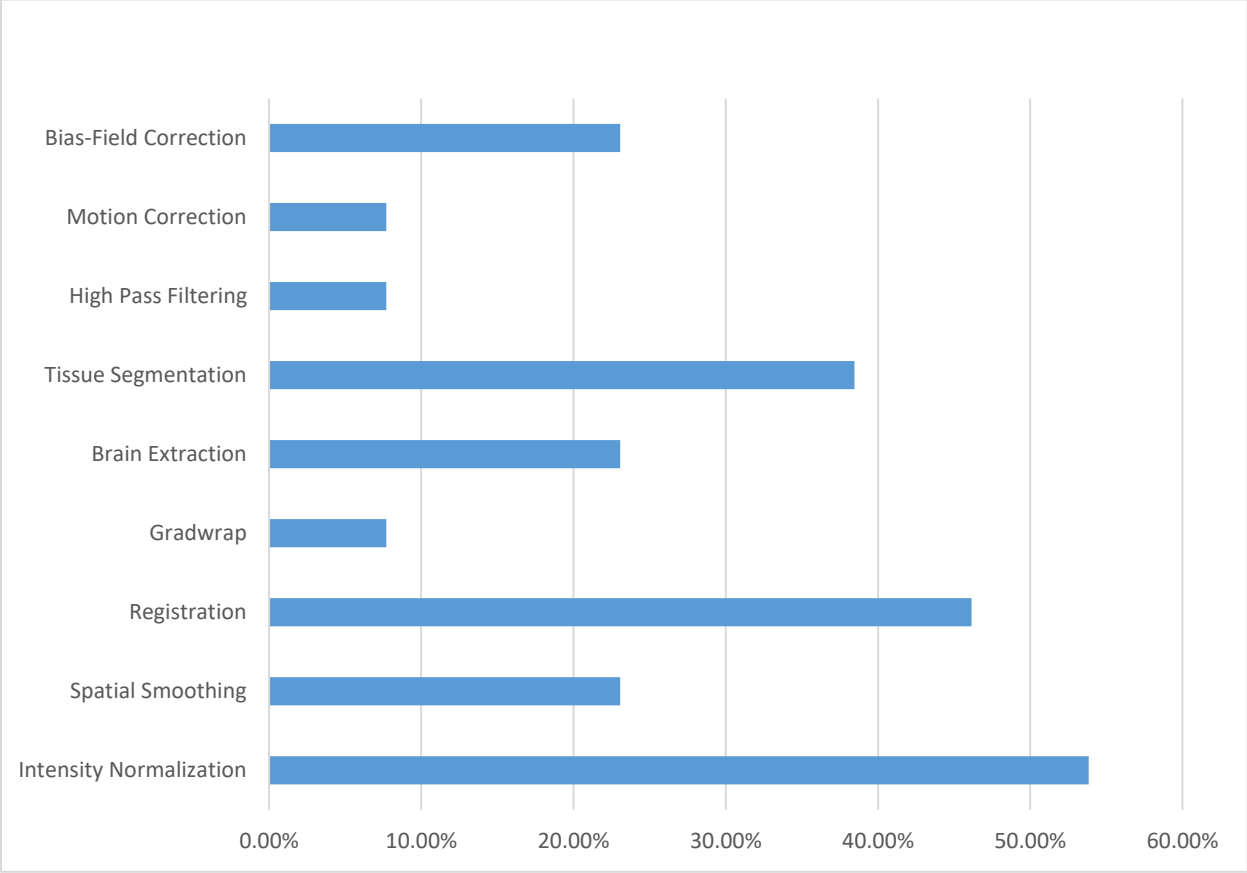


Figure 2-6 Utilization percentage of preprocessing techniques

Based on the derived characteristics from neuroimaging modalities**,** the handling of input data may be divided into four distinct categories.

Voxel-based input management: In this technique, the voxel intensity values of whole neuroimaging modalities or tissue components of modalities are used. It requires a registration procedure to map all the images into a standard 3D space [50]. Basaia et al. [140] used tissue segmentation and registration of MRI scans to create probability maps of GM, WM, and CSF tissue in MNI space. Using the DARTEL approach, Lin et al.[143] segmented the MRI scans into GM, WM, and CSF and then utilized PCA and SFS to precisely choose the best features. In several research articles [139], [141], [142],authors did the complete analysis of neuroimaging modalities. In [145], the authors conducted a multimodal study of the whole brain (18 FDG-PET and AV-45 PET) and co-registration utilizing the AC–PC line. The benefit of a comprehensive brain analysis is that spatial information is fully incorporated; however, high dimensionality and processing effort are drawbacks of this strategy.

Slice-base input management: Using either the researcher's reasoning or conventional projections, such as the horizontal plane, frontal plane, and median plane, 2D slices are derived from 3D scans [50]. All the data from 3D brain scans cannot be turned into a 2D slice, this prevents a comprehensive analysis of the brain. Ramzan et al. [137] transform rs-fMRI scans into 2D slices along the image height and time axes and get 6160 2D slices per fMRI scan. In [19] Wu et al. transform each scan into 160 2D slices. Yang et al. [146]averaged 65 slices along the Z-axis for each scan. In [147] , Hon et al. selected the 32 most useful slices from the axial plane view of each scan using an entropy-based sorting technique. During the training and testing phase of the model, the number of parameters may be reduced from millions to hundreds using this method. Due to the use of 2D representations of 3D scans, the spatial dependence between consecutive 2D pictures is lost.

ROI-based input management: For early diagnosis of AD, the informative components impacted in the early stages of the disease have been used[50], [142]. The hippocampus, amygdala, entorhinal cortex, grey matter, and temporal and parietal lobes are analyzed for categorization of AD or MCI[169], [170], while the amygdala and hippocampus are utilized to predict the progression of MCI to AD [171]. This method also needs prior knowledge of the brain atlas, such as from automated anatomical labeling (AAL) or Kabbani's reference work [172], [173]. Abrol et

al.[22] found the most discriminative brain areas for categorizing pMCI vs. sMCI by estimating occlusion sensitivity using the network occlusion technique [174] .

Lu et al.[144] segmented the sMRI into GM and WM and split the GM into 85 cortical and subcortical anatomical ROIs. Li et al. [138] segmented the hippocampus from other areas and created a binary mask for each hippocampus. After calculating the centroid of each hippocampus, patches were taken from the centroid for further implementation. The ROI-based method has the benefit of minimal feature dimensions and simple interpretation, but it disregards any abnormality specifics.

Patch-Based Input Management: In this method, disease-related patterns are found by extracting the features of patches or 3D cubes[50], [142]. The main challenge is to convey both patch-level and image-level properties by using the most discriminative patches [109]. The segmented ROI is divided into smaller, patches in [144]. Li et al. [138] segmented hippocampus to form a binary mask, which was then cropped at the centroid to produce 3D patches. This method has the advantage of being sensitive to minute changes, but it is still difficult to determine which patch will provide the most useful information.

The main points are summed up below.

- A 2D-slice input offers the benefit of fewer training parameters and a less sophisticated network, but the disadvantage of losing spatial dependence between adjacent slices.
- Voxel-based input management takes all brain information into account but treats all areas identically, disregarding their biological structure, and has the disadvantage of high dimensionality.
- An ROI-based input is easily interpretable, however, the absence of tiny aberrations in the ROI might cause harm to discriminative information and limit the real relevance of retrieved features. In **Appendix A.2**, you can find a summary of the preprocessing pipelines used by all the articles in this SLR.

The percentage of input management techniques used is depicted in **Figure 2-7**.

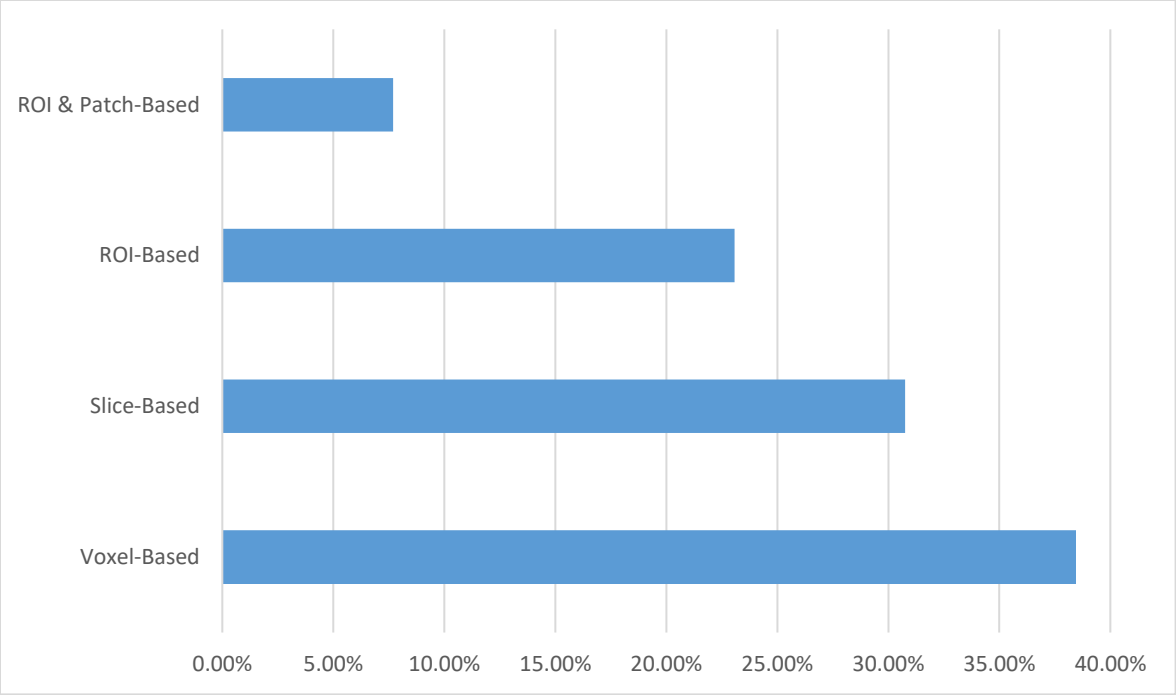


Figure 2-7 Utilization percentage of input management techniques

2.4.6 RQ4. Current level of accuracy and other performance metrics

Two of the thirteen articles used multiclass classification, while eleven utilized binary classification. All of them used the accuracy performance measure uniformly. Ramzan et al. [137] classified multiclass (AD, NC, sMCI, pMCI, SMC, and MCI) with an average accuracy of 97.92%. Using GoogleNet and CaffeNet, Wu et al. [19] conducted three-way discrimination between the NC, sMCI, and pMCI and got an overall accuracy score of 83.23 and 87.78%, respectively. **Table 2-3** summarizes the achieved accuracy, DL models with TL approaches, and input management techniques with types of biomarkers.

The maximum accuracy for classifying sMCI vs. pMCI obtained was 87.78% by utilizing a CNN-based, pre-trained CaffeNet model and slice-based input management with the sMRI biomarker by Wu et al.[19]; however, this accuracy was attained during the multiclass classification. Using 3D CNN, local TL, and multimodalities (18F-FDG PET and AV-45 PET) with voxel-based input management, Choi et al.[145] obtained 96.0% and 84.2% accuracy for binary classification of AD vs. CN and sMCI vs. pMCI, respectively. Using 3D CNN with local TL and sMRI with voxel-based input management, Basaia et al. [140] achieved 98.20% accuracy for the AD vs. CN challenge. In future research, the accuracy of sMCI vs. pMCI task needs to be improved.

DL and TL combination	Input Management	AD vs. CN	pMCI vs. sMCI	Ref.
CNN Based ResNet-18	Slice Based & rs-fMRI	97.92% [Multiclass]		[137]
3D CNN & local TL	Voxel Based & sMRI	98.20%	74.90%	[140]
3D CNN & CAE & local TL	Voxel Based & sMRI	85.24% ± 3.97%	73.23% ± 4.21%	[141]
		86.60% ± 3.66%	73.95% ± 4.82%	
CNN & local TL	Voxel Based & sMRI	85.00% ± 4.00%	73.00%± 5.00%	[142]
	ROI Based & sMRI	88.00% ± 3.00%	78.00%±7.00%	
	Patch Based & sMRI	83.00% ± 2.00%	77.00%±4.00%	
	Slice Based & sMRI	79.00% ± 4.00%	-	
CNN based Deep ResNet	ROI Based & sMRI	91.30%	77.80%	[22]
CNN based CaffeNet	Slice Based & sMRI	87.78% [Multiclass]		[19]
CNN based GoogleNet	Slice Based & sMRI	83.23% [Multiclass]		
CNN based CaffeNet	Slice Based & sMRI	-	77.98%	[143]
DNN & SAE for local TL	ROI-Patch Based & sMRI +18F-FDG PET	93.58%	82.50%	[144]
3D CNN & local TL	Voxel Based & 18F-FDG PET + AV-45 PET	96.00%	84.20%	[145]
CNN based AlexNet	Slice Based & 18F-FDG	-	72.19%	[146]
CNN based VGG16 (E2EL)	Slice Based & sMRI	74.12%	-	[147]
CNN based VGG16 (TL)		92.30%	-	
CNN based InceptionV4(TL)		96.25%	-	
DenseNets & RNN	ROI-Patch Based & sMRI	89.10%	72.50%	[138]
3D CNN & local TL	Voxel Based & sMRI	-	76.00%	[139]

Table 2-3 Accuracy attained by using the DL model in conjunction with the TL strategy, an input management strategy, and a neuroimaging biomarker.

2.4.7 RQ5. Datasets of Neuroimaging Biomarkers

There are several online datasets of neuroimaging biomarkers that have been created with the goal of making neuroimaging data freely accessible to the scientific community to aid in future discoveries in the neurosciences. Online resources for open data sets include ADNI[23], OASIS[24], AIBL[25], IXI[26], and MIRIAD[27].

ADNI: It was publicized in 2004 as a USD 60 million, 5-year cooperation initiative, headquartered in North America, by the National Institute on Aging (NIA) and the National Institute of Biomedical Imaging and Bioengineering (NIBIB).

As of now, there are three different cohorts:

- **ADNI-1:** (400 MCI, 200 sMCI, and 200 CN) MRI and FDG-PET biomarkers.
- **ADNI-GO (Grand Opportunities):** The existing ADNI-1 cohort + 200 early MCI subjects.
- **ADNI-2** is a continuation of ADNI-1, the ADNI-GO cohort + (150 CN, 100 early MCI, 150 mild AD and 150 late MCI), that began in 2016 and will continue until 2022.

92% of articles utilize this dataset alone or in conjunction with other datasets, making it the most frequently used dataset.

OASIS: The most recent edition of the OASIS series is OASIS-4. It includes the neuroimaging data obtained from 1098 patients, aged between 43 and 95 years, who participated in 2168 MRI sessions and 1608 PET sessions (609 CN and 489 persons with various degrees of cognitive disability). In 16% of the articles, it was utilized.

AIBL: This adds scientific value to the ADNI cohort and is referred to as the Australian ADNI. It includes 50 MRI/PiB[28] scans, 200 MRI/florbetapir[29] scans, and 250 MR/flutemetamol scans. In 8% of the articles, it was utilized.

IXI: There are 600 scans of CN subjects. In 8% of the articles, it has been used.

[23] https://adni.loni.usc.edu/
[24] https://www.oasis-brains.org/
[25] https://aibl.csiro.au/
[26] https://brain-development.org/ixi-dataset/
[27] https://www.nitrc.org/projects/miriad/

[28] radiotracer Pittsburgh Compound B (PiB)
[29] 18-F amyloid radiotracers (florbetapir and flutemetamol)

MIRIAD: 46 AD and 23 CN MRI scans are included in the MIRIAD database. None of the included articles have utilized it.

Several research, however, favored use their own datasets. Basaia et al. [140] created a self-regulating dataset called "MILAN" of 3D T1W MRI scans that were collected from 229 participants (124 AD, 50 MCI, and 55 CN). **Appendix A.1** contains a summary of all additional information, including sample size and the combination of various datasets from all included articles.

2.4.8 RQ6. Integrated Development Environments

A variety of software packages are available to aid researchers in the preprocessing of neuroimaging data and the deployment of DL and TL approaches.

For the preprocessing of neuroimaging biomarkers, the following packages were used:

- FSL[30]: This is the complete toolkit for processing data from sMRI, fMRI, and DTI neuroimaging. It has been used in [137], [140].
- SPM12[31]: This program is designed to analyze neuroimaging data sequences, which may consist of a series of images from several angles. [22], [141], [143] made use of it.
- Nipype[32]: It is an open-source Python project created by the Nipy[33] community, which dedicated to the analysis of neuroimaging data using the Python programming language. It provides a consistent interface to existing brain imaging software packages, such as AFNI[34], ANTS[35], Camino[36], FreeSurfer[37] , FSL, MNE[38] , Slicer[39], and SPM, and makes it easier for them to communicate with one another within a single flow. It has been utilized in [142] .
- MATLAB[40]: The image processing toolbox allows for the automation of standard image processing procedures. The researcher may segment image data, handle big datasets in

[30] https://fsl.fmrib.ox.ac.uk/fsl/fslwiki
[31] https://www.fil.ion.ucl.ac.uk/spm/software/spm12/
[32] https://nipype.readthedocs.io/en/latest/
[33] https://nipy.org/
[34] https://afni.nimh.nih.gov/
[35] https://www.nitrc.org/projects/ants
[36] https://www.nitrc.org/projects/camino/
[37] https://surfer.nmr.mgh.harvard.edu/
[38] https://mne.tools/dev/auto_tutorials/inverse/90_phantom_4DBTi.html
[39] https://www.slicer.org/
[40] https://www.mathworks.com/solutions/neuroscience.html

batches, register photos, generate histograms, and work with ROIs. [19], [147] made use of it.

- NifTi Toolkit[41] [139]: This is a tool for processing and interpreting neuroimaging data that can be acquired using the packages MRIcron[42] . In has been utilized in [146].
- FreeSurfer: The Laboratory for Computational Neuroimaging USA created this open-source collection for studying and manipulating MRI images. It has been utilized in[144].
- MRIcron: This program gives format headers for exporting neuroimaging biomarkers to other platforms and allows the users to look at medical pictures in different formats. Neuroimaging data may be efficiently viewed and exported, and ROIs can be found using it. It has been utilized in [139].

The authors of [145] did not preprocess the FDG and AV45 PET scans.

DL and TL algorithms have been implemented using software frameworks such as CAFFE[43] in [19], [137], [146], Keras[44] in [138], [147], Theano[45] in [140], Pytorch[46] in [22], [142], MatConvNet[47] in [145], and the Deep Learning Toolbox[48] in [144] . The specifics of the software package were not covered in three articles[139], [141], [143].

2.4.9 RQ7. Managing over fitting in Deep Learning models

The effectiveness of classifiers on an unseen test dataset is significantly affected by the size of the dataset [106]. There are just a few hundred samples available in the existing datasets of people with AD and MCI [50], [102]. Because DL algorithms have an enormous number of training parameters, they are easily overfit when trained on a smaller set of data[8], [41], [102].

Overfitting has been minimized using the methods described below in the included articles.

- **Data Augmentation:** This is a method for boosting the diversity of training datasets without resorting to further data collection [175], [176]. It takes the current data and creates

[41] https://www.mathworks.com/matlabcentral/fileexchange/8797-tools-for-nifti-and-analyze-image
[42] https://www.nitrc.org/projects/mricron

[43] https://caffe.berkeleyvision.org/
[44] https://keras.io/about/
[45] https://theano-pymc.readthedocs.io/en/latest/
[46] https://pytorch.org/
[47] https://www.vlfeat.org/matconvnet/
[48] https://www.mathworks.com/products/deep-learning.html

new data samples. There are two distinct approaches to this. The first type is **Transformation techniques**, which include a wide variety of straightforward changes to the training data, such as translation, rotation, reflection, distortion, blurring, flipping, cropping, noise injection, gamma correction, scaling, and intensity variations via arbitrary adjustments to brightness, contrast, saturation, and hue [45], [175]–[177]. In [140], [141], transformation techniques were applied. Second, **Neuroimaging data synthesis** [178] is technique for creating a new dataset with similarities to the original dataset. To apply this to neuroimaging biomarkers, generative adversarial networks (GANs) [41], [179] might be used. However, further research is needed in this field, and the efficiency of the syntesis pictures in predicting AD has not yet been shown.

In order to produce sample image patches from MRI, Wu et al.[19] used a new approach for data augmentation that was based on Shin's [180] image integration technique. Choi et al.[145] enhanced PET pictures by swiping them from left to right. To extract additional patches, Li et al.[138] shifts the hippocampus centroid by 2 voxels in the x and z dimensions.

By giving DL models a better weight initialization, TL approaches may lessen overfitting. It also cuts down on the period required to train the DL models.

- **Regularization:** Dropout, weight decay are a few methods that may be used [10], [89]. Dropout [49] is based on the principle of randomly and independently removing neurons, setting their output value to zero, and reducing the complexity and overfitting risk of the network. It increases the DL model's capacity for generalization.

  By regularizing the new weights and multiplying them by a factor lower than one, weight decay[50] also makes the model more generalizable and less complicated. Several of the studies that were considered included dropout and weight [22], [139], [142], [144], [147] .

- **Batch normalization [51]:** It is used to normalize the input to a layer for each mini batch while training DL models. It expedites model training and improves model performance and generalizability.

[49] https://www.analyticsvidhya.com/blog/2022/08/dropout-regularization-in-deep-learning/
[50] https://programmathically.com/weight-decay-in-neural-networks/
[51] https://www.analyticsvidhya.com/blog/2021/03/introduction-to-batch-normalization/

- **Early stopping [52]:** This entails halting the training procedure earlier. It aids in figuring out how many training iterations are necessary before the model becomes significantly overfit. It has been used in [22], [144].
- **Cross-Validation [53] :** This approach involves repeatedly splitting the database into training and testing sets while maintaining the same proportions and rotating the instances. It is a statistical technique for evaluating classifier performance. The k-fold cross validation technique is the most popular. In our analysis, k's value falls between 5 and 10. About 85% of the included articles used it.

2.4.10 RQ8. Research opportunities, constraints, and limitations

DL models with TL approaches have shown better results; however, the following issues still need to be addressed.

- Diagnosing the change from MCI to AD is more important because it lets patients and their loved one's plan and helps the doctors figure out which patients need the right treatments. So far, the best accuracy for this task has been 84.20% [145] , which needs to be improved.
- Multimodality approach performs better then single modality[144], [145]. In the future, new combinations of biomarkers may be used to predict the onset of AD.
- The dearth of substantial neuroimaging datasets causes generalizability issues, despite the use of TL and augmentation. To make a generalizable classifier, large datasets can be created using neuroimaging data synthesis
- It may be challenging to show precisely which features have been retrieved and to regulate how those features affect the inference and relative prominence of other characteristics [13]. As a result, it is challenging to eliminate any biases that the input datasets may have generated. Filter visualization and activation maps[54] were used by [144] to figure out what factors had a big effect on the result.

[52] https://www.analyticsvidhya.com/blog/2021/03/introduction-to-batch-normalization/

[53] https://www.baeldung.com/cs/k-fold-cross-validation

[54] How to Visualize Filters and Feature Maps in Convolutional Neural Networks - MachineLearningMastery.com

2.5 Conclusion

We began our review by discussing the concepts of AD and MCI, followed by a discussion of current recommendations for early detection of AD by using MRI, fMRI, and PET neuroimaging biomarkers. These biomarkers can be combined with some other parameters, such as genetic information, MMSE scores, and Mini-Cog Test findings, for improved categorization.

Neuroimaging methods are advised to undergo the least amount of pre-processing (intensity normalization, registration, and brain extraction). For enhanced feature extraction, multimodality might be applied. Most of the **research made use of CNNs with various TL methods**. However, **certain very well-known TL designs that have been very helpful in the analysis of medical data, like EfficienNet and DenseNet [24], [181], have not been implemented for AD.**

Several issues need to be rectified, including overfitting in relation to the use of small datasets and reproducibility in respect to the randomness occurring during training. Furthermore, it might be challenging to decide which features are more important for a reliable decision assistance system.

DL models required a lengthy training period because of the vast amount of 3D input and random weight initialization. Researchers utilized local TL with 3D CNN networks in studies that reported high accuracy**. The most efficient blends of DL and TL with the right biomarkers may be found to further enhance these findings**.

Because of these limitations, it is likely that AI-based systems that help doctors to make decisions about the current state of AD are not available. The primary objective of this dissertation is to assist in overcoming such limitations by proposing the following: (1) the identification and implementation of the most efficient DL model using E2EL (**see Chapter 3**); and (2) the improvement of the accuracy of the identified model through the utilization of the fusion of E2EL and TL **(see Chapter 4**).

Chapter 3: Implementation of eight 3D CNN architectures based on E2EL by using T1W MRI scans

Summary: As shown in Figure 1-1**, the content of this chapter is about the second contribution, "Identification and implementation of the optimum DL model**," and it is intended to help achieve the second goal, "to provide the neuroradiologist with the computer-interpretable information they need to analyze an MRI scan". We employ sMRI scans to create 3D CNN models using E2EL and a voxel-based input management strategy for two categorization tasks: AD vs. CN and AD vs. sMCI.

The MP-RAGE T1 MRI images of 245 AD patients and 229 with sMCI were obtained from the ADNI dataset. The IXI dataset was used to obtain the 245 T1 MRI scans of CN individuals. Denoising, N4 bias field correction, extraction of the brain, and registration were the four procedures that comprised the preprocessing of every scan.

A total of eight different 3D CNN-based architectures are developed, tested, and compared. These comprised **DenseNet121, DenseNet169, DenseNet201, and DenseNet264, as well as EfficientNet-B0 through B3.** Since these models have already been proven successful in other types of medical image analysis, **the AD domain has not yet been investigated by them.**

The comparison study was carried out with the assistance of **a Ranking Mechanism, Comprehensive Indicators, and an Efficiency-Effectiveness Graph. The DenseNet264 performed exceptionally well** in both types of classification, achieving an accuracy of 82.5% and an area under the curve (AUC) of 87.63% during training and an accuracy of 81.03% during testing in relation to the AD vs. sMCI. It also achieved 100.00% accuracy and 100.00% AUC during training and 99.66% accuracy when comparing AD to CN. On the other hand, the **EfficientNet-B0 had the greatest overall metrics and the lowest number of model parameters for the training of the sMCI against the AD task in the Efficiency-Effects Graph**.

3.1 Introduction

E2EL design is the fundamental tenet of DL. E2EL key advantage is that it simultaneously improves every stage of the processing pipeline, potentially leading to optimum results [25], [182].

For neuroimaging analysis, Oh et al.[141] proposed an E2E hierarchical structure with levels ranging from 1 to 4. At **level 1**, feature selection and extraction are done manually. To be

used as input into DL networks, 3D data is reconfigured into 1D vector form .In **level 2**, 3D scans are either split up into WM, GM, CSF, or hippocampal areas or turned into 2D slices before being used as input for a DL model. Preprocessed 3D volume data are utilized as the **level 3** input for DL models. A 3D MR image acquired from a scanner is fed directly into DL networks at **level 4**. Nevertheless, the author is unaware of any studies that have used this level and documented it in the literature.

In April 2022, the author used Sysrev[55] and the search queries "stable MCI deep learning" or "stable MCI vs. Alzheimer by using transfer learning" or "early detection of Alzheimer by using deep learning" or "stable MCI vs. AD classification for early detection of AD" to identify ongoing research status at levels 1, 2, 3, and 4 and discovered 121 articles, 81 of which were considered and 40 were not. The 81 articles were analyzed and split into levels afterwards.

The preponderance of published studies used Level 1 [85], [87], [183]–[187] or Level 2 [15]– [20], [22], [86], [106], [135], [136], [142], [144], [145], [169], [170], [189]–[212], [212], [213] .The outcomes are dependent on specific software, hyperparameter tweaking, and manual noise reduction. Performance assessments in these research articles only used a subset of the original datasets due to these interdependencies, avoiding obvious outliers and making a fair performance comparison difficult [50], [76].There were just a few research that used the Level 3 hierarchy[133], [135], [140], [141], [213], [214].

Author decided to use level 3 E2EL, MRI scans, and a voxel-based input management approach to identify the most effective 3D CNN model for categorizing phases of AD . This decision was based on the context described above as well as the recommendations and limitations discussed in Chapter 2.The methodology is discussed in Section 2, along with the datasets, the preprocessing of the MRI scans, 3D CNNs, experimental setup and algorithm. In Section 3, the results are shown. Section 3 also includes a comparative analysis of all models using a ranking system, comprehensive performance indicators, and efficiency effect graphs. We also compared our results with the results of the published state-of-the-art methodologies.

[55] https://sysrev.com/u/2642/p/118998

3.2 Materials and methods

Figure 3-1 depicts the block diagram of the proposed technique. There are four steps to get there.

1. Use the ADNI and IXI datasets to get T1W MRI scans of people with AD, CN, and sMCI.
2. Preprocess the data using AntsPyNet and split it into two parts; the first set will be used for training and validation, while the second set will be used for testing.
3. Using E2EL and MONAI, build eight cutting-edge classifiers that are based on 3D CNNs.
4. A comparison study was carried out with the goal of locating the most efficient 3D CNN model. Metrics such as accuracy, area under the curve (AUC), precision, recall, and F1-score were used, along with ranking mechanisms, comprehensive performance indicators, and efficiency impact graphs.

3.2.1 Participants

Datasets from ADNI and IXI were utilized. 719 MP-RAGE T1W MRI scans were acquired in Neuroimaging Informatics Technology Initiative (NIfTI) format as AD (245 scans), CN (245), and sMCI (229). AD and sMCI were downloaded from ADNI, whereas CN scans were downloaded from IXI. Only MCI scans that were defined in the ADNI description files to be stable for at least 4 years and up to 15 years were downloaded.

Neuroimaging file types, called NIfTI[56], are often used in image analytics in neuroscience and neuroradiology research. It offers the advantage of storing a volume in a single file that has a header and raw data after that. It enables rapid loading and processing [50].

We present the technique here to identify patients with sMCI. The following two CSV files must be examined to identify the years of stability of the MCI state:

1. **ADNIMERGE:** Can be downloaded from Study data > Test data > Data for challenges[57]. By looking at the values in the columns DX.bl (baseline diagnosis) and DX (status), you can identify if MCI is stable, if it has turned into AD, or if it has become normal.
2. **Diagnosis Summary** :Downloaded from Study Data > Assessments > ALL Diagnosis[58]. To distinguish between the different stages of AD, DXCHANGE column

[56] NIfTI (file format) | Radiology Reference Article | Radiopaedia.org
[57] https://ida.loni.usc.edu/pages/access/studyData.jsp?categoryId=43&subCategoryId=94
[58] https://ida.loni.usc.edu/pages/access/studyData.jsp?categoryId=12&subCategoryId=37

value can be analyzed. It describes the participant's change in cognitive state between the prior appointment and the present one.

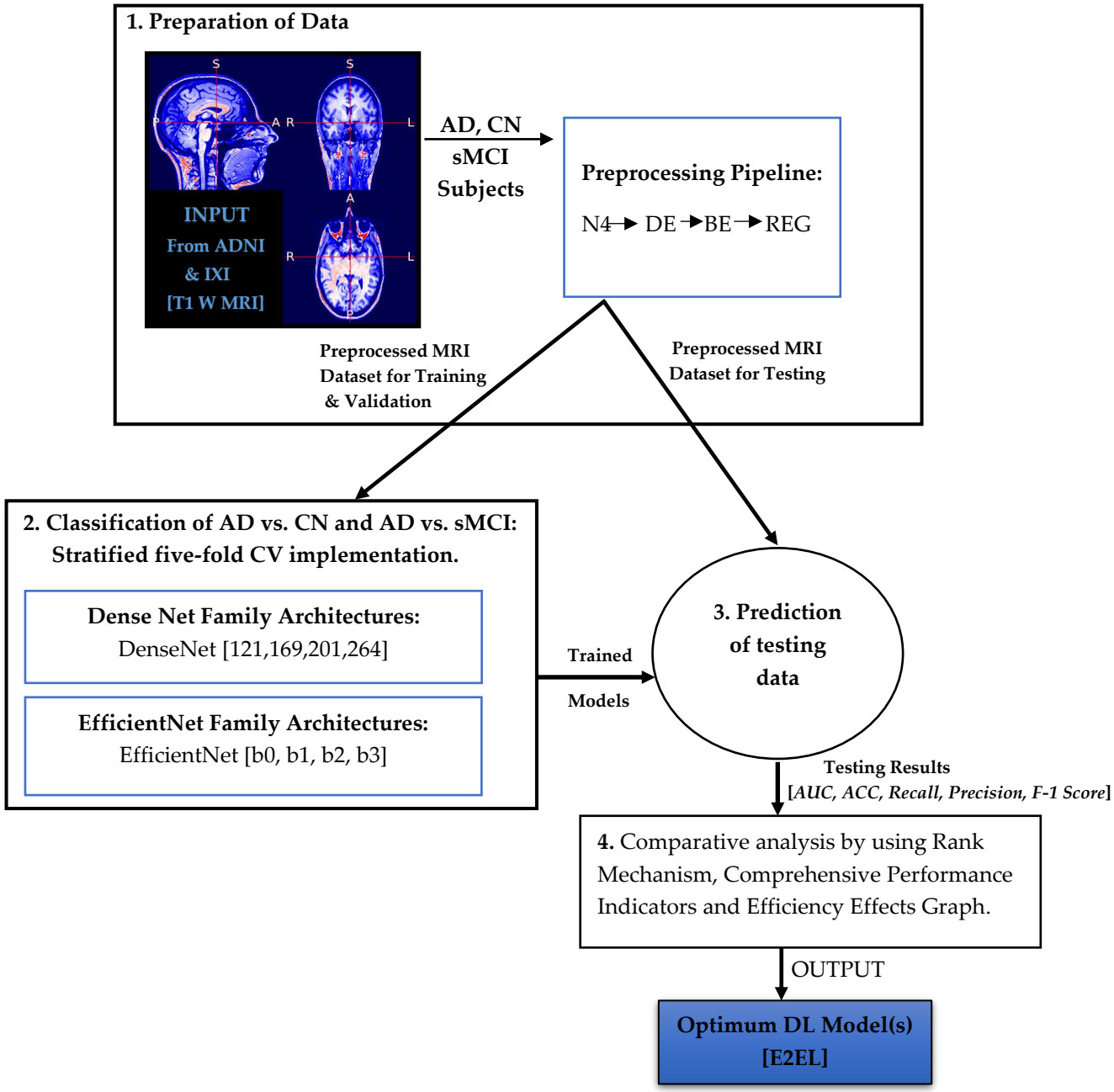


Figure 3-1: Proposed Methodology N4: N4 bias field correction; DE: Denoising; BE: Brain Extraction; REG: Registration

The author carefully studied around 6000 rows of the csv data. Only 229 patients with sMCI have clear records of each visit and had remained stable for at least 4 years up to 15 years. Other column values such as Age, Gender, the clinical dementia rating score (CDR) [216], MMSE score, and the

4 allele of apolipoprotein E (APOE4) [217], [218] were all ignored since the focus of our research was E2EL/Level 3. MRI scans downloaded from ADNI often had 256x256x176 voxels with 1 mm x 1 mm x 1.2 mm sizes, but those from IXI commonly had 256x256x256 voxels with 1 mm x 1mm x 1 mm sizes.

3.2.2 Preprocessing Pipeline

Using the ANtsPyNet[59] package, the Advanced Normalization Tools pipeline [215]–[217] was applied to each scan, as shown in **Algorithm 1**. ANtsPyNet is a collection of Python-based DL architectures and applications for basic medical image processing. The following methods are included in the preprocessing pipeline:

1. A popular method for rectifying low-frequency intensity non-uniformity, sometimes referred to as a "bias" or "gain field," in MRI scans is the **N4 bias field correction** technique [168]. This approach utilizes a simple parametric model without tissue categorization. The ants.utils.n4 bias field correction( )[60] function was used to perform it, which was then followed by denoising.
2. **Denoising** [61] primary goal is to estimate the original image by reducing noise in an image. Image noise may cause by a wide range of internal and external variables, many of which are challenging to avoid in practical circumstances. In image classification task, where retrieving the original image information is essential for successful results, image denoising is thus important.

   We perform denoising in two steps: first, we add salt-and-pepper noise [62] to the MRI scans, and then we use the utility ants. denoise image( ) [63], to eliminate noise by using a spatially adaptive filter [218]. The procedure is then followed by the brain extraction.
3. **Brain extraction** performed to remove non-brain tissues such as those in the neck and skull by using ANtsPyNet brain extraction ( ) utility [64], which employs a pretrained 3D U-net model brainy[65]. The ability of brainy to leverage contextual information between slices is its primary advantage. Brainy could estimate the brain mask for a volume of

[59] https://pypi.org/project/antspynet/
[60] ants.utils.bias_correction — ANTsPy master documentation
[61] http://stanford.edu/class/ee367/Winter2016/Chaudhari_Report.pdf
[62] https://www.simonwenkel.com/notes/ai/practical/vision/progressive-sprinkles-and-salt-and-pepper-noise.html
[63] https://antspy.readthedocs.io/en/latest/_modules/ants/utils/denoise_image.html
[64] https://antsx.github.io/ANTsPyNet/docs/build/html/utilities.html#antspynet.utilities.brain_extraction
[65] brainy/brainy at master · neuronets/brainy · GitHub

256x256x256 in three seconds, regardless of orientation. In our implementation, predicting the brain mask of each image took around five seconds. The procedure is then followed by the registration.

4. The purpose of **registration** is to remove spatial differences across participants and reduce the impact of transformations relative to a reference orientation.

   We performed affine fast registration [219] using the ANTsPy utility ants.registration ( ) [66] in the MNI 152 template[67]. MNI 152 is a template for a universal brain atlas that was generated by averaging 152 structural scans into one large image using nonlinear registration. Affine registration aligns the mean, linear shape, and orientation of data without requiring a problem-specific adjustment.

Algorithm 1 MRI Preprocessing Pipeline

Require: Baseline T1 weighted MRI scans
Ensure: Pre-processed MRI scans .

```
Image ← ants.image_read (path)                    ▷ Input
step 2 for bias field correction
Image ← ants.utils.n4_bias_field_correction (Image)
step 3 and 4 for denoising by using salt and pepper noise
Image_t ← Image + np.random.randn (*Image.shape).astype ('float32')*5
Image ← ants.denoise_image (Image_t, ants.get_mask (Image))
step 5 and 6 for brain extraction
Image_t ← antspynet.utilities.brain_extraction (Image)
Image ← Image*Image_t)
step 7 and 8 for registration in MNI 152 template
Image_t ← ants.registration (fixed=MNI152Template, moving=Image, AffineFast)
Image ← ants.apply_transforms (fixed=MNI152, Image_t ['fwdtransforms']))
```

Algorithm 1 : Preprocessing pipeline of MRI scans.

The output of each step for a CN subject is shown in **Figure 3-2**, along with the dimension, spacing, and origin. After preprocessing, the dimensions changed from 256x256x15 to **182x218x182.**

[66] https://antspy.readthedocs.io/en/latest/registration.html
[67] https://www.lead-dbs.org/about-the-mni-spaces/

The origin of the anatomical coordinate system is the location of the first voxel, whereas the spacing represents the distance between voxels along each axis.

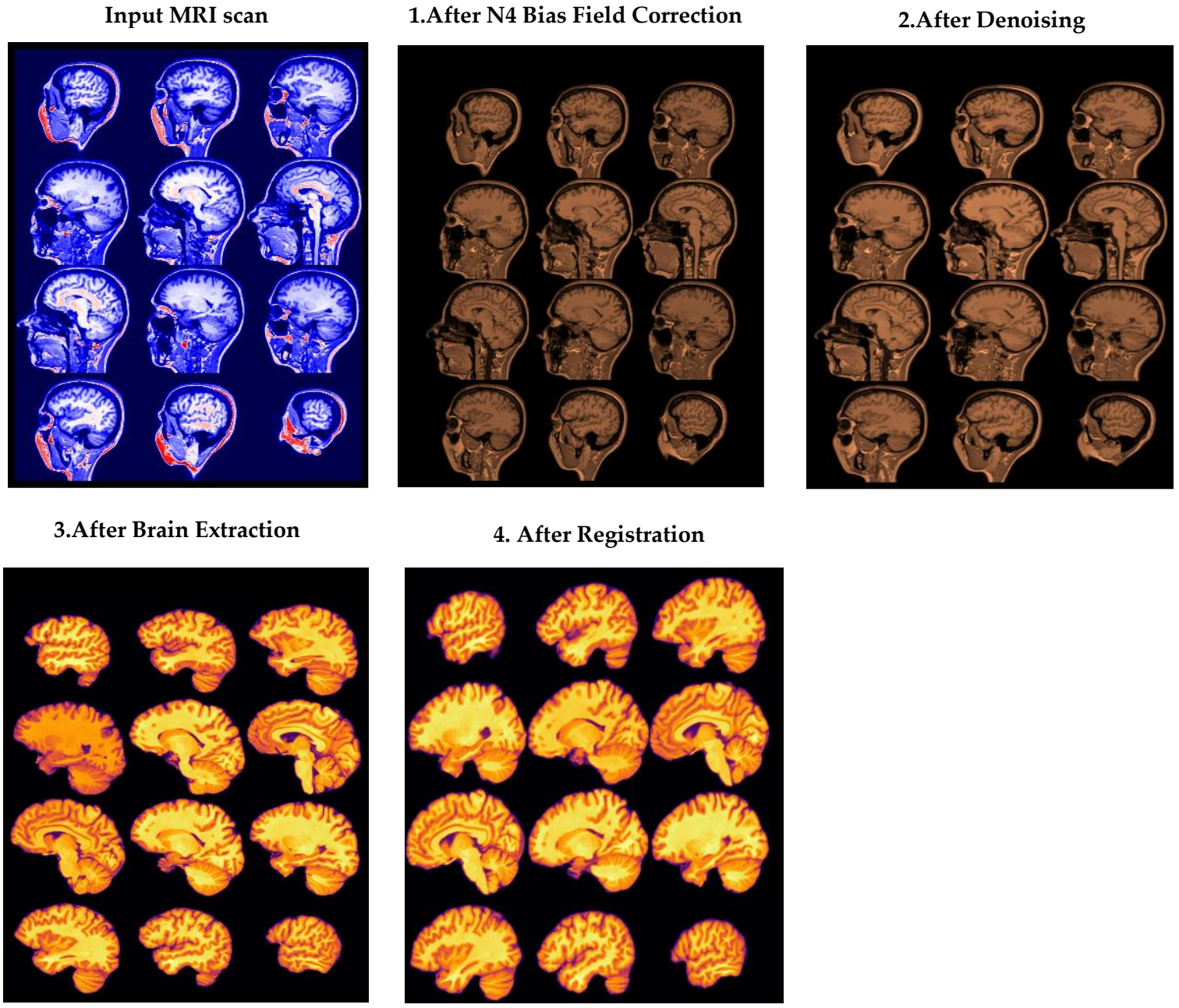


Figure 3-2 Output of each stage of the preprocessing pipeline for a cognitively normal subject

{ 1->2->3->4 }

Input scan : Dimension: (256, 256, 150) Spacing : (0.9375, 0.9375, 1.2) Origin : (88.6399, -116.532, -112.1136)

Output scan :Dimension : (182, 218, 182) Spacing : (1.0, 1.0, 1.0) Origin : (-90.0, 126.0, -72.0)

The MNI152 template standard also altered the spacing and origin values for all MRI scans to (1.0, 1.0, 1.0) and (-90.0, 126.0, -72.0) accordingly. The pre-processing for a single MRI scan took around two minutes. In **Appendix B**, a link to the Python script for Algorithm 1 is has been given.

3.2.3 Implemented 3D CNN

A variety of cutting-edge, 3D-CNN-based designs, particularly those from the DenseNet [181] and EfficientNet [220] families, have lately shown excellent performance in a variety of medical data categorization tasks [24], [44], [45], [86], [221]–[223], but we were unable to find many implementations of these models to detect AD. Based on this context, using Level 3/E2EL, the subsequent 3D CNN models for classifying AD vs. CN and AD vs. sMCI individuals were implemented.

- A feed-forward neural network called **DenseNet** connects every layer to the one before it. DenseNet has L(L + 1)/2 direct connections, whereas traditional L-layer convolutional networks have L connections [76]. DenseNets provide several enticing benefits, including the ability to resolve the "vanishing gradient" problem, boost feature reuse, and significantly reduce the number of parameters. We have implemented four 3D DenseNet designs, each of which has four Dense Blocks with varying layer counts. **Table 3-1** displays the size, number of parameters, and number of layers in each block of the deployed designs.

DenseNet (Version)	Parameters	Layers block wise	Size
121	11.2 M	[6,12,24,16]	9393 MB
169	18.5 M	[6,12,32,32]	9892 MB
201	25.3 M	[6,12,48,32]	10924 MB
264	40.2 M	[6,12,64,48]	12424 MB

Table 3-1 Employed structural design of DenseNet.

- **EfficientNet** [220] is a lightweight model built on the AutoML[68] framework. It was used to create a baseline EfficientNet-B0 network, and uniformly scale up the depth, width, and

[68] AutoML for large scale image classification and object detection – Google AI Blog (googleblog.com)

resolutions using a simpler and more effective compound coefficient to improve EfficientNet models B1–B7.

These models performed well and outperformed the prior CNN models on the ImageNet datasets [23] . EfficientNets are smaller, faster, and more generalizable, resulting in better accuracy. It's common to utilize them for TL, although they only assist with 2D data.

We have implemented EfficientNetB0-B3 by using E2EL. The architecture specifics are shown in **Table 3-2**. B4-B7 could not be employed in the present research because to the model's direct 3D volume input, the increasing amount of parameters, and the constrained GPU and Memory resources.

EfficientNet	Parameters	Size
B0	4.6 M	7801 MB
B1	7.4 M	10223 MB
B2	8.7 M	10631 MB
B3	12.01 M	14294 MB

Table 3-2 Employed structural design of EfficientNet.

3.2.4 Exploratory Setup :

For the binary classification task, AD vs. CN and sMCI vs. AD, eight 3D CNN architectures were implemented by using stratified 5-fold cross-validation (CV) [69] as shown in Algorithm 2.

The number of groups into which a certain data sample is to be divided is indicated by the procedure's only parameter, k. The process is hence often referred to as k-fold CV. When a particular value for k is selected, it may be substituted for k in the reference to the model, such as when k=5 is used to refer to 5-fold CV.

CV is a well-liked technique since it is easy to comprehend and often yields a less biased or too optimistic assessment of the model ability than other techniques, including a straightforward train/test split. Data folding may be managed by rules like guaranteeing that each fold has the same number of observations with a certain category value, such as the class result value. This process is known as stratified CV[70].

[69] https://machinelearningmastery.com/k-fold-cross-validation/
[70] https://machinelearningmastery.com/cross-validation-for-imbalanced-classification/

3.2.4.1 Frameworks, Tools, and IDEs

The whole implementation was completed using **Google Colab Pro+**[71], which was made available to the public in August 2021. Some of its most notable features include the ability to run in the background, early access to more powerful GPUs, and increased memory availability. Asynchronous data loading and multiprocessing are facilitated by GPUs.

Despite not guaranteeing compatibility with a specific GPU, Colab Pro+ does offer priority on the available options. Even with Pro+, GPU quality might decline after periods of intensive use. Pro+ offers Tesla V100 or P100 NVIDIA DL GPU[72] with CUDA support.

The "High-RAM" option of Colab runtime met its objective by providing **52.8 GB RAM**. Runtime support is supposed to be 24 hours as stated in Colab's specs[73], yet we only got assistance for a maximum of 8 hours. For this reason, we could not run all the folds at the same time in this setup, as finishing a fold with 50 epochs requires approximately 2 hours.

To implement 5-fold, stratified CV, we must create five data sets (DATASETS 1-5) for training and validation, with the same class ratio as the original dataset across all folds by using self-written code in Python. In Appendix B, a link to the code and datasets is given. According to the preprocessing pipeline discussed in Algorithm 1, all scans were processed. After preprocessing, datasets [1–5] were created and utilized in folds 1–5, respectively. To train and validate Fold [n], we utilized Dataset [n].

All models in this research were developed using Medical Open Network for Artificial Intelligence (MONAI) [74], a publicly accessible, community-supported, PyTorch-based toolkit for deep learning in medical imaging. It offers domain-optimized core capabilities for developing healthcare imaging training workflows inside a native PyTorch paradigm. The goal of Project MONAI, a collaborative effort between NVIDIA and King's College London, is to bring together academic and industrial researchers working on artificial intelligence in the field of medical imaging to develop and disseminate best practices. The author was unable to locate any research that used MONAI to identify AD.

[71] https://towardsdatascience.com/google-colab-pro-is-it-worth-49-99-c542770b8e56
[72] https://www.nvidia.com/en-us/data-center/v100/
[73] https://blog.paperspace.com/alternative-to-google-colab-pro/
[74] https://monai.io/

3.2.4.2 Network Hyperparameters

Hyperparameters are an integral part in training neural networks[75]. A hyperparameter regulates the learning process, therefore its values have an immediate effect on other model parameters, like as weights and biases, and subsequently on the model's performance.

Two critical hyperparameters, learning rate and number of epochs, were tuned. We trained and tested the model on three different combinations of learning rate and number of epochs for the AD vs. sMCI task and based on the comparison of their performances as shown in Fig. 3-3, the learning rate was set to 0.0001 and the number of epochs was set to 50. The link to the Python scripts for tuning has given in Appendix B. The batch size has been adjusted to 2 because of the limited available RAM.

The Adam optimizer [224] has been utilized. Adam was the first "adaptive optimizer" to gain massive acceptance [225]. Rather of using a separate learning rate scheduler, adaptive optimizers integrate learning rate optimization directly into the optimizer itself. Even further, Adam controls the learning rates on a weight-based scale. That is, it provides a learning rate for each independent variable in the model. The value that Adam assigns to this learning rate is a feature of the optimizer's implementation that cannot be modified directly[76]. Learning rate schedulers like ReduceLROnPlateau and EarlyStopping are unnecessary because of the way Adam is implemented.

In addition, the ROCAUC Metric and the cross-entropy loss function were used. A neural network's output value during the backpropagation process is often a minimum, which is far lower than the real target value. It is often difficult for the neural network to use the supplied data to update its weights and enhance itself when the gradient is quite modest [77].

The cross-entropy function's logarithm aids the network in detecting and correcting such little faults. CNN may utilize this change as guidance in the intended direction far more efficiently with the cross-entropy function than with the mean-squared error function. The equation of the cross-entropy loss function is $L_{CE} = -\sum_1^n t_i \log(p_i)$ for n classes.

[75] https://www.analyticsvidhya.com/blog/2022/05/impact-of-hyperparameters-on-a-deep-learning-model/
[76] Gentle Introduction to the Adam Optimization Algorithm for Deep Learning - MachineLearningMastery.com
[77] https://www.superdatascience.com/blogs/convolutional-neural-networks-cnn-softmax-crossentropy

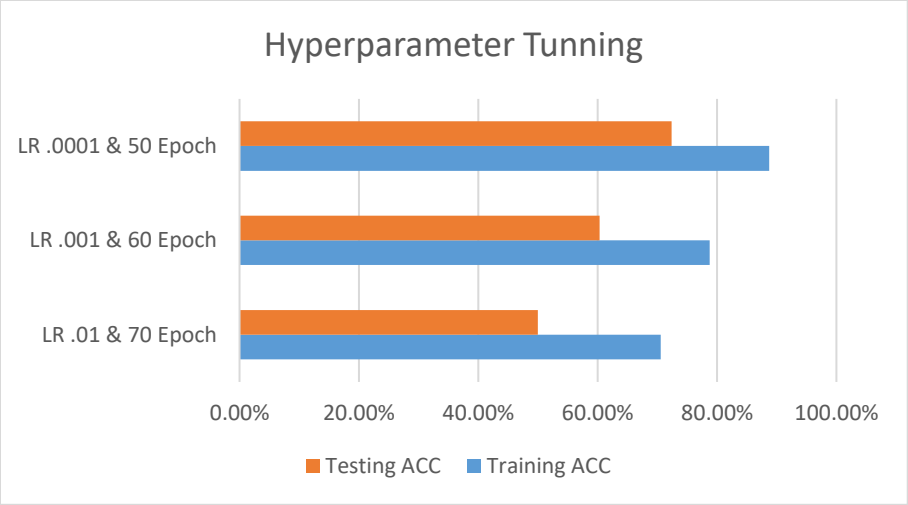


Figure 3-3 Tuning of learning rate and number of epochs

The ROCAUC Metric measures a model's ability to distinguish between classes. The higher the AUC, the better the model predicts 0 classes as 0 and 1 class as 1.

3.2.4.3 Evaluation Matrices

The following measures were used to assess the results since they give vital information for a comprehensive assessment of models.

- **Confusion Matrix:** A confusion matrix, as shown in **Table 3-3**, which is an X-by-X grid where X is the number of class labels, can be used to figure out how well a classification model works[78]. Four outputs are expected from the confusion matrix: True positive (TP), True negative (TN), False positive (FP), and False negative (FN). If a patient with AD predicts within the AD group, he or she is classified as a TP, and if not, as a FN. Similarly, TN represents the number of correctly predicted normal patients, and FP represents the number of incorrectly predicted normal patients.

	Positive Predictions	Negative Prediction
Real Positive	TP	FN
Real Negative	FP	TN

Table 3-3 Confusion Matrix

[78] https://www.sciencedirect.com/topics/engineering/confusion-matrix

- The following five metrics [79] were used to assess the results:
    - **Accuracy (ACC)** = (TP + TN) / (TP + TN + FN + FP).
    - **Precision** = TP / (TP + FP), it represents how effectively the model learns about the positive sample features and is specific to the predicted result. The better the precision, the more precise the positive sample prediction.
    - **Recall** = TP/ (TP + FN), the greater the recall rate, the more accurately the target sample can be anticipated, and the less likely a faulty sample will be overlooked.
    - The **F1-score** = $\frac{2 \cdot TP}{2 \cdot TP + FP + FN}$, is the harmonic mean of precision and recall, and it assesses a test's accuracy. In general, precision and recall are in conflict; consequently, the F1-score is presented as a composite metric to balance the impacts of precision and recall and more accurately assess classifiers.
    - **ROC-AUC**, receiver operating characteristic-area under the curve (ROCAUC). It's a graph for determining how effectively a model can discriminate between two classes. The larger the area under the ROC curve, the better the classification method. When you look at the ROC curve, you can see the tradeoff between sensitivity and specificity for all conceivable thresholds, not just the one that the modeling approach picked.

3.2.4.4 Algorithm

Algorithm 2 presents the implemented procedure for the classification tasks. As described in Section 3.2.4.1, MRI scans from DATASET [C] were used for training and validation in fold C. After the preprocessing of all data, the MRI scans were distributed as follows:

- For AD vs. CN task :
    - AD: 160 for training, 40 for validation, 45 for testing.
    - CN: 160 for training, 40 for validation, 45 for testing.
- For AD vs. sMCI task :
    - AD: 160 for training, 40 for validation, 29 for testing.
    - sMCI: 160 for training, 40 for validation, 29 for testing.

[79] https://blog.paperspace.com/deep-learning-metrics-precision-recall-accuracy/

Algorithm 2 For AD vs. CN and AD vs. sMCI classification task

Require: Pre-processed T1 MP-RAGE MRI scans of AD and sMCI and CN Subjects
Ensure: Trained and validated 3D CNN model for improved inference .

batchsize ← 2,*workers* ← 2 ▷ ***Parameters for Data Loaders***
trainloader ← *DataLoader*(*trainingdataset*[1])
validationloader ← *DataLoader*(*validationdataset*[1])
testingloader ← *DataLoader*(*testingdataset*)
device ← *cuda* ▷ *Parameters for the model*
spatial_dims ← 3 ▷ for 3D Input
input_channels ← 1 ▷ gray-scale images
num_classes ← 2 ▷ Number of output classes
optimizer ← *Adam*, *learning_rate* ← .0001 ▷ Tuned Hyper-parameter
loss_function ← **CrossEntropyLoss()**
auc_metric ← **ROCAUCMetric()**
model ← **EfficientNet(B0,B1,B2,B3) and DenseNet(121,169,201,264)**
NO_epoch ← 50,*C* ← 1 ▷ Tuned Hyper-parameter
while *C* ≤ 5 **do** ▷ Outer loop for carrying out 5-fold stratified CV
    epoch ← 1 ,*best_metric* ← −1 ▷ *begin a PyTorch training*
    while *epoch* ≤ *NO_epoch* **do** ▷ Training for the 100 epoch
        epoch_loss ← 0
        model.train() ▷ sets the mode to train.
        for `batch_data in trainloader` **do** ▷ employ mini-batches
            inputs&*labels* ← *batch_data*
            optimizer.zero_grad() ▷ Gradients should be set to 0.
            outputs ← *model*(*inputs*) ▷ The model infers.
            loss ← *loss_function*(*outputs*, *labels*) ▷ Determine the loss.
            loss.backward() ▷ compute gradients during the backward pass.
            optimizer.step() ▷ update the learning weights.
            epoch_loss ← *epoch_loss*+*loss*
        end for
        model.eval() with torch.no_grad() ▷ sets the model to evaluate
        for `val_data in validationloader` **do** ▷ evaluation loop
            val_inputs&*val_labels* ← *val_data*
            val_pred ← *model*(*val_inputs*)
            val_loss ← *loss_function*(*val_pred*, *val_labels*)
        end for
        Determine the acc_metric() and auc_metric() values
        if *acc_metric* ≥best_metric **then** ▷ Save the best metric model.
            A ← *val_labels* , *B* ← *val_pred* , *best_metric* ← *acc_metric*
            torch.save(model.state_dict())
        end if
        epoch ←epoch+1
    end while ▷ Training Completed for FOLD C.
    draw_confusion_matrix(A,B),draw_roc(A,B)
    model.eval() with torch.no_grad() ▷ sets the model to test
    for `test_data in testingloader` **do** ▷ test using new MRI scans
        test_inputs&*test_labels* ← *test_data*
        test_pred ← *model*(*test_inputs*)
    end for
    draw_confusion_matrix(test_pred,test_labels),draw_roc(test_pred,test_labels)
    trainloader ← *DataLoader*(*trainingdataset*[*C* + 1])
    validationloader ← *DataLoader*(*validationdataset*[*C* + 1])
    C ←C+1 ▷ reset the model; End to End Learning.
end while=0

Algorithm 2 : Used for implementing eight CNN models for AD vs. CN and AD vs. sMCI tasks

As indicated in steps 2 and 3, we utilized MRI scans from DATASET[1] for training and validation. We initialized the training and validation loaders from the DATSET [C] MRI scans to be analyzed in fold C in steps 48 and 49. In steps 43 to 46, we used the model with the best weights from each fold to assess unseen MRI scans. The model was reset to apply E2EL after each fold. In step 12, we have mentioned that this algorithm will be used for implementing all eight models. We executed 40 Python scripts for eight models, each with five folds. In Appendix B, a link to all the scripts and their findings is provided.

3.3 Results

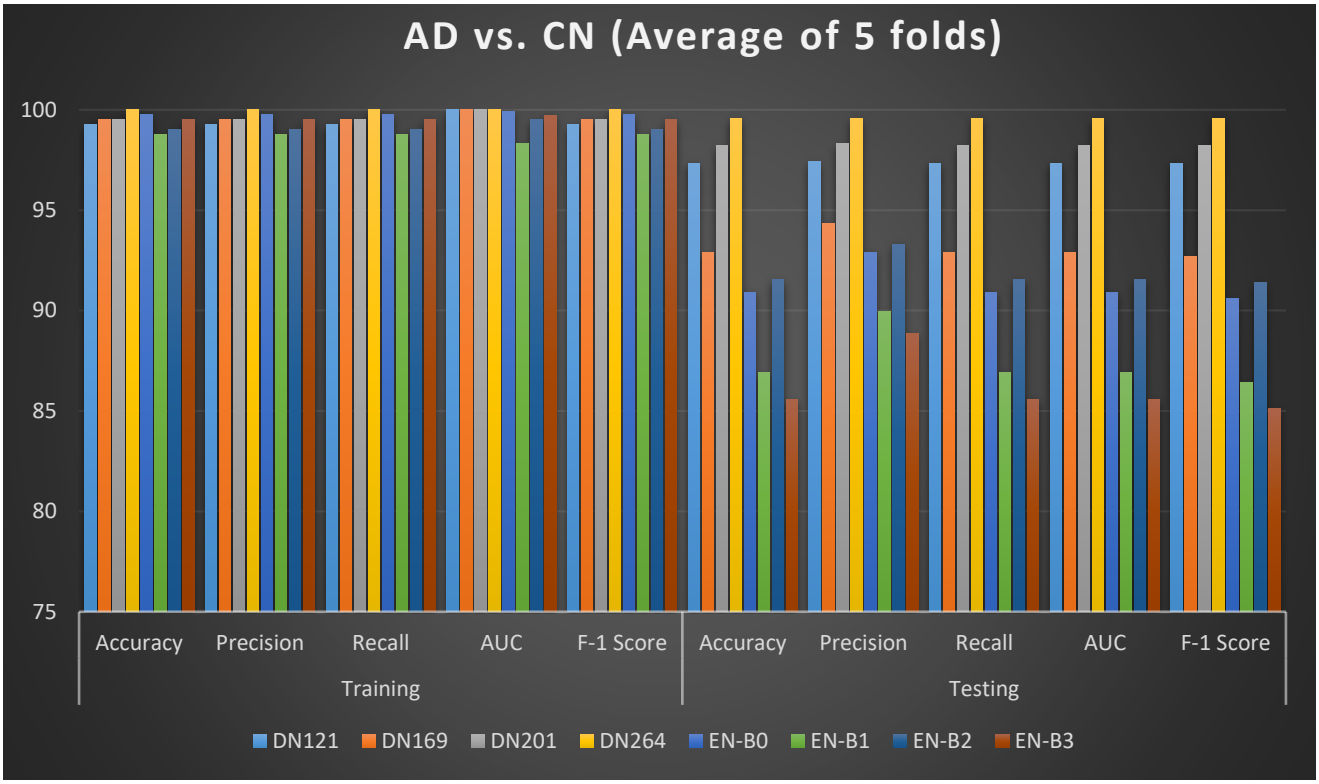
Table 3-4 and Figure 3-4 show the results of five metrics (precision, recall, F1-score, accuracy, and AUC) used to compare the eight models for both training and testing. During testing, the DenseNet-based models beat the EfficientNet-based models for both classifications by a margin of 7 to 14 % for the AD vs. CN classification and by a margin of 5 to 10 % for the AD vs. sMCI classification. In both classification methods, DenseNet 264 outperformed the rest of the DenseNet family. EfficientNet-B0 beat the other models based on EfficientNet. During the AD vs. sMCI testing, DenseNet201 and EfficientNet-B0 did better than DenseNet264 in a few evaluation matrices by a small margin of 1% to 3%.

Figures 3-5 to 3-8 depicts the confusion matrix and ROCAUC for the best fold of DenseNet264 for both classification tasks. The confusion matrix and ROCAUC for other models for each fold can be seen by using the link given in **Appendix B**.

Validation after training was done on a total of 80 MRI scans, with 40 scans used for each class for the AD vs. sMCI task. **Figures 3–5** show that the TP for sMCI patients was 35 and the FN was 5, while the TN for AD patients was 37 and the FP was 3. Each class achieved an AUC of 90%.

58 MRI scans were utilized for testing, including 29 scans from subjects with AD and 29 from subjects with sMCI, which had not been used during the training or validation procedures. **Figure 3–6** shows that those with AD had a TN of 27 and an FP of 2, while those with sMCI had a TP of 24 and a FN of 5. The AUC was 87.9% for both classes.

Due to the paucity of information in the ADNI database indicating the state of each participant after x years of stability and the fact that their anatomical features are almost identical to those of AD scans, only a small percentage of sMCI and AD individuals were incorrectly classified.



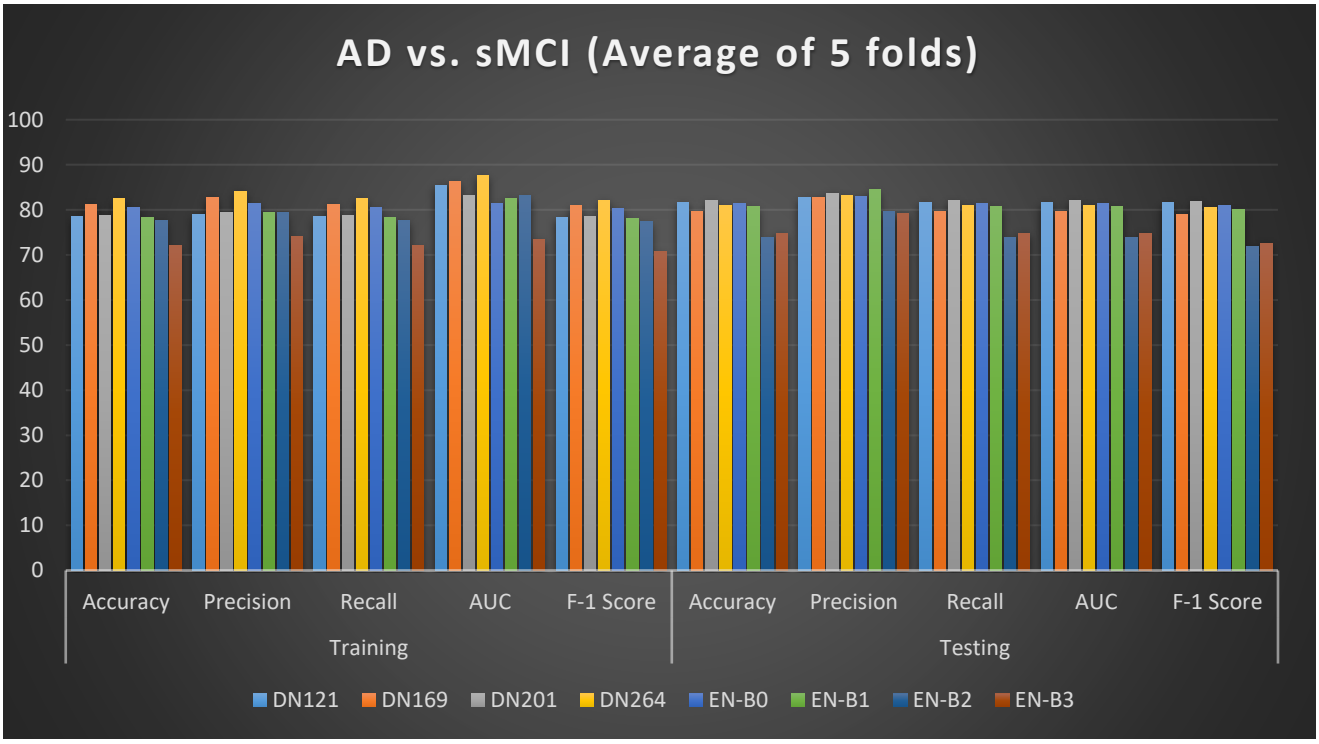


Figure 3-4 DN: DenseNet; EN: EfficientNet; Graphical Results of five assessment matrices of eight DL modes for the AD vs. CN and AD vs. sMCI tasks

AD vs. CN	Matrices	DN121	DN169	DN201	DN264	EN-B0	EN-B1	EN-B2	EN-B3
TRAIN	ACC	99.25	99.50	99.50	**100**	99.75	98.75	99.00	99.50
	Precision	99.28	99.52	99.52	**100**	99.75	98.78	99.01	99.51
	Recall	99.25	99.50	99.50	**100**	99.75	98.75	99.00	99.50
	AUC	100	100	100	**100**	99.94	98.34	99.52	99.74
	F-1 Sc.	99.25	99.50	99.50	**100**	99.75	98.75	99.00	99.50
TEST	Accuracy	97.33	92.89	98.22	**99.55**	90.91	86.91	91.55	85.56
	Precision	97.41	94.35	98.30	**99.56**	92.90	89.98	93.30	88.84
	Recall	97.33	92.91	98.22	**99.55**	90.91	86.93	91.55	85.58
	AUC	97.33	92.89	98.22	**99.55**	90.91	86.91	91.55	85.56
	F-1 Score	97.33	92.674	98.22	**99.55**	90.58	86.40	91.38	85.11

AD vs. sMCI	Matrices	DN121	DN169	DN201	DN264	EN-B0	EN-B1	EN-B2	EN-B3
TRAIN	Accuracy	78.50	81.25	78.75	**82.50**	80.50	78.25	77.75	72.00
	Precision	78.99	82.83	79.45	**84.10**	81.44	79.39	79.33	73.99
	Recall	78.48	81.25	78.75	**82.50**	80.50	78.25	77.75	72.00
	AUC	85.42	86.23	83.20	**87.63**	81.38	82.59	83.14	73.49
	F-1 Score	78.42	80.94	78.63	**82.15**	80.33	78.06	77.38	70.71
TEST	Accuracy	81.72	79.65	82.06	**81.03**	81.38	80.69	73.79	74.83
	Precision	82.72	82.83	83.70	**83.29**	82.91	84.49	79.65	79.22
	Recall	81.72	79.65	82.06	**81.03**	81.38	80.69	73.80	74.83
	AUC	81.73	79.65	82.06	**81.03**	81.38	80.69	73.79	74.83
	F-1 Score	81.59	79.07	81.84	**80.60**	80.96	80.00	71.97	72.60

Table 3-4 Tabular Results of five assessment matrices of eight DL modes for the AD vs. CN and AD vs. sMCI tasks (average of a 5- fold stratified CV) [101]

Validation after training was done on a total of 80 MRI scans, with 40 scans used for each class for the AD vs. CN task .Figures 3–7 show that the TP for sMCI patients was 40 and the FN was 0, the TN for AD patients was also 40 and the FP was 0. Each class achieved an AUC of 100%.

90 MRI scans were utilized for testing, including 45 scans from subjects with AD and 45 from subjects with CN, which had not been used during the training or validation procedures. Figure 3–8 shows that those with AD had a TN of 45 and an FP of 0, while those with CN also had a TP of 45 and a FN of 0. The AUC was 100% for both classes. DenseNet264 achieved an average ACC of 99.55% for the AD vs. CN task, however obtained a maximum accuracy of 82.50% for AD vs. sMCI classification, which must be improved.

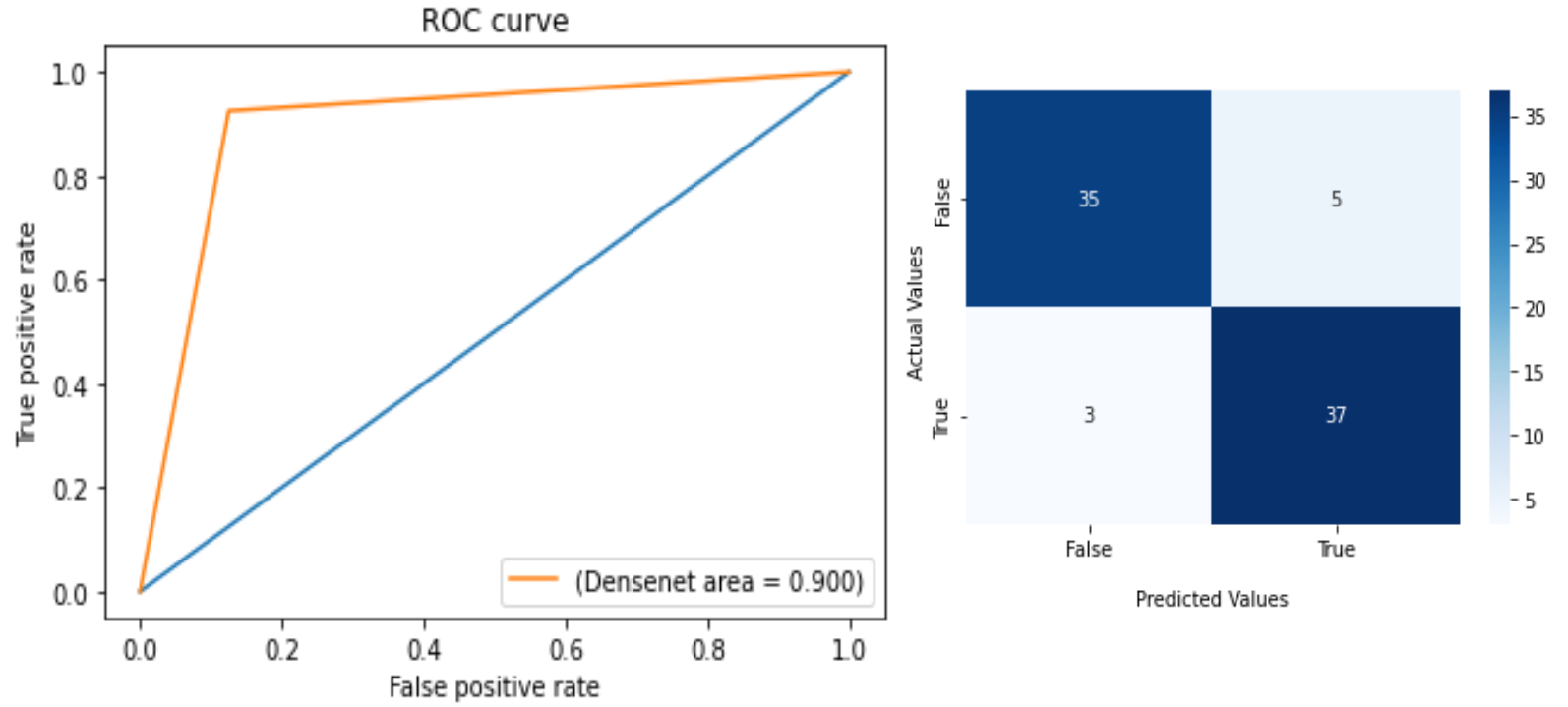


Figure 3-5 ROC-AUC curve and confusion matrix for the best fold of training for the AD vs. sMCI task

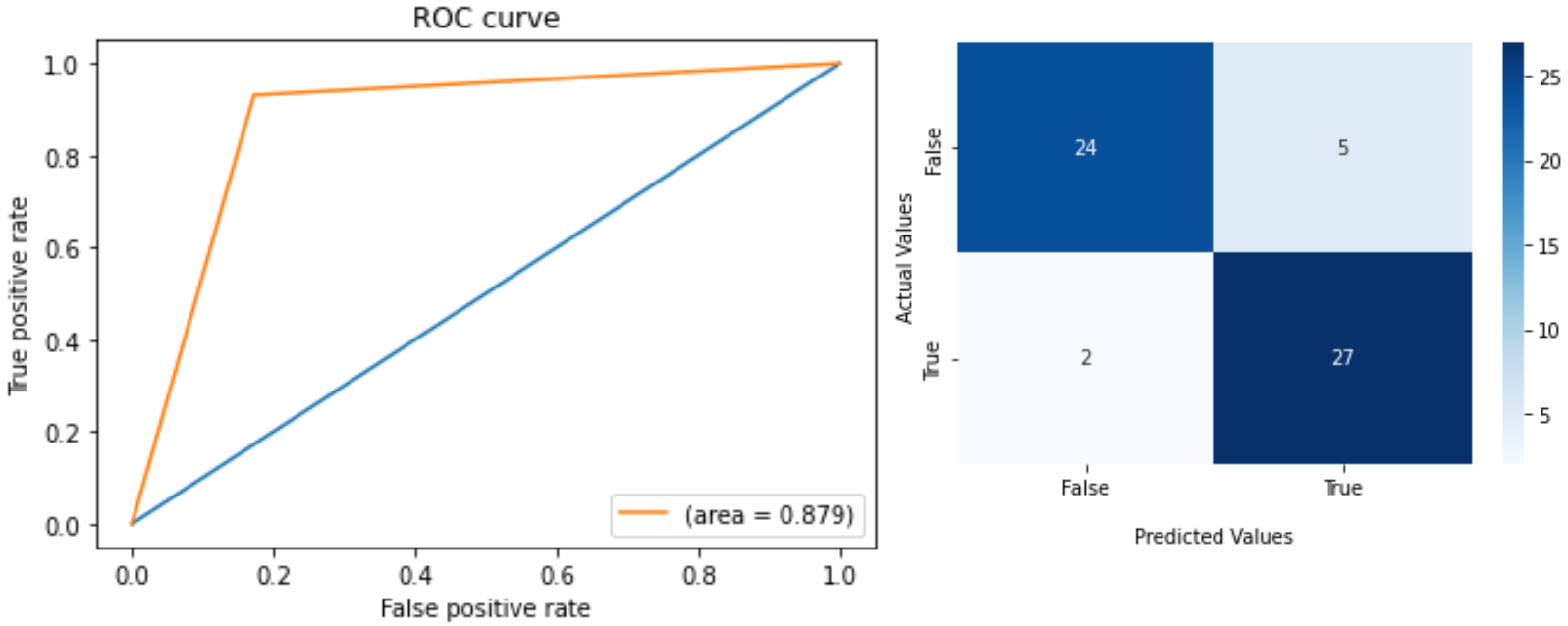


Figure 3-6 ROC-AUC curve and confusion matrix for the best fold of testing for the AD vs. sMCI task

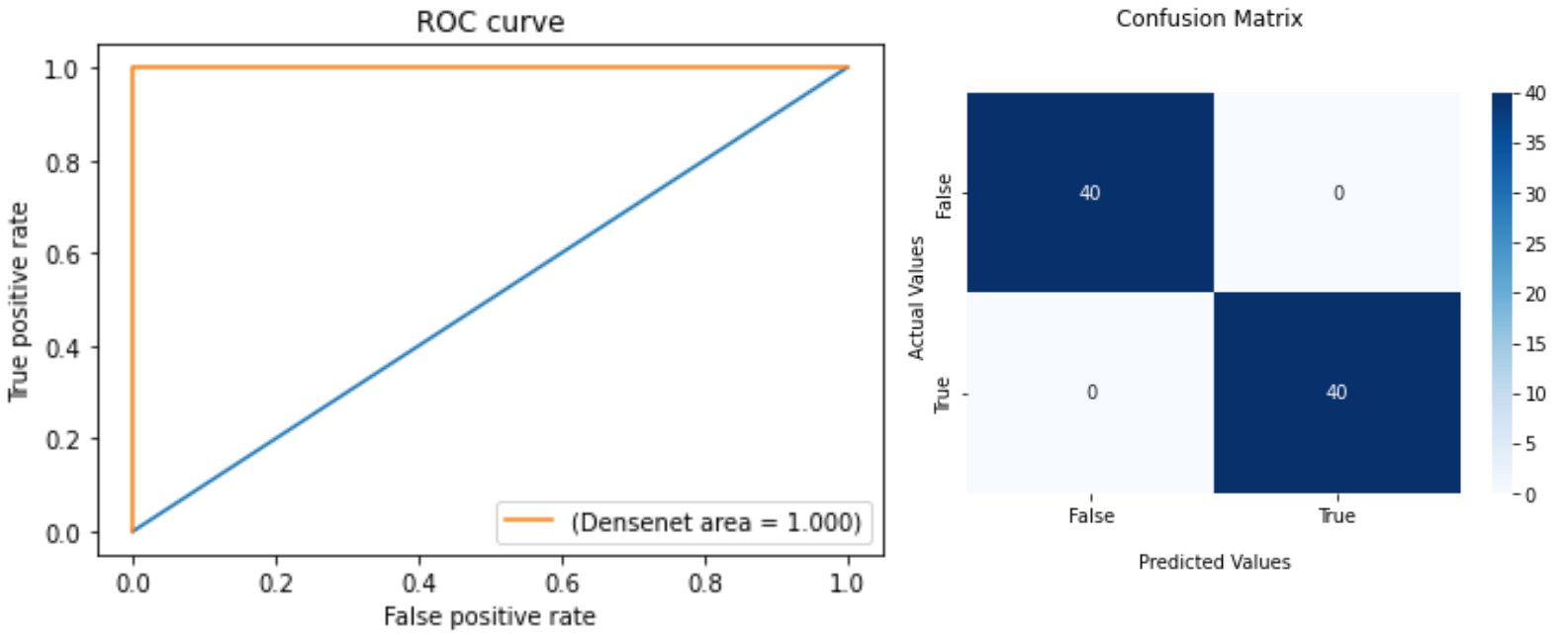


Figure 3-7 ROC-AUC curve and confusion matrix for the best fold of training for the AD vs. CN task

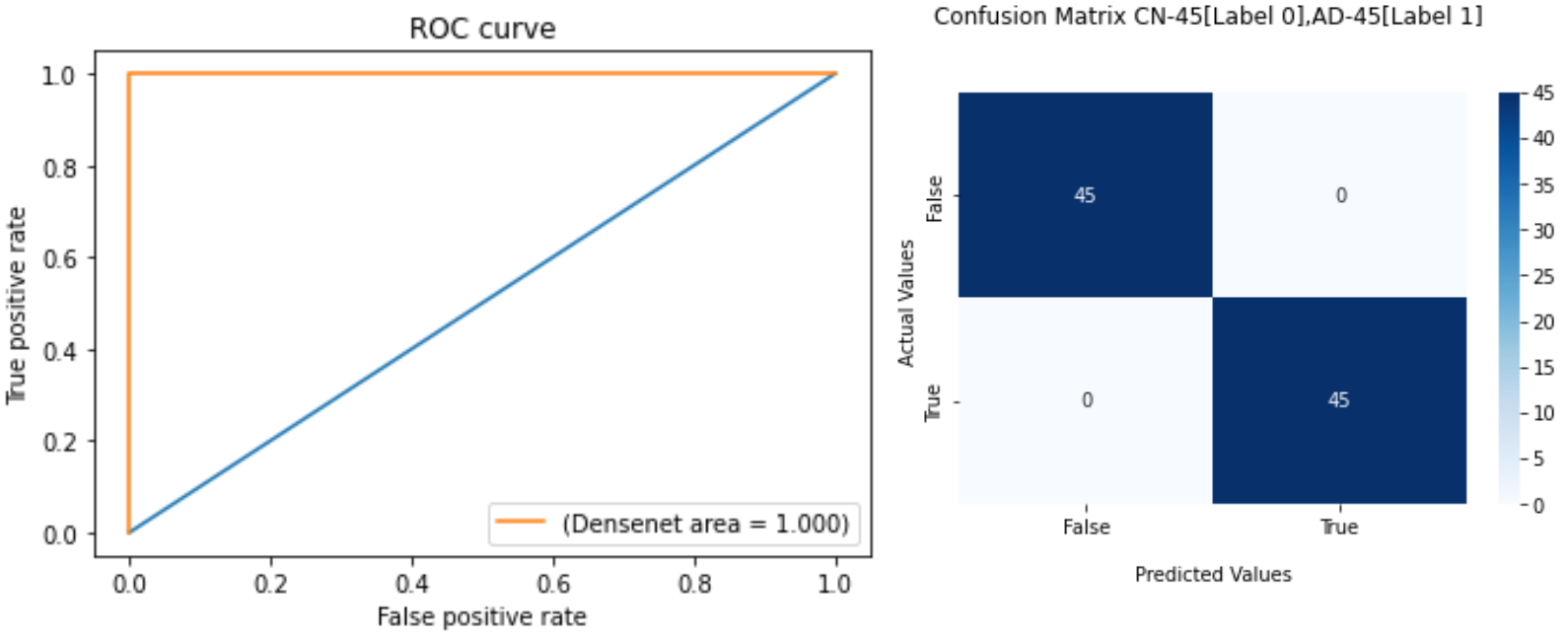


Figure 3-8 ROC-AUC curve and confusion matrix for the best fold of testing for the AD vs. CN task

3.4 Comparative Analysis

The results of the relative comparison of the implemented models were described in detail using the following two techniques:

3.4.1 Ranking Mechanism

The ranking method described by Zorlu et al. [226] was used. Individually, for the training and testing, the total rank of every model was determined. Because there were eight models and the highest rating was awarded to the model with the best performance, the maximum rating for each performance index was set at eight. After that, each model's overall performance rating was calculated by summing the ranks of its training and test datasets.

DenseNet264 achieved the top rating among the eight DL models for both binary classification tasks as shown in Table **3-5 and 3-6**.

DL Model	Evaluation Phase	ACC Rank	Precision Rank	Recall Rank	AUC Rank	F-1 Score Rank	Total Rank	Grand Total Rank
DN121	**TRAIN**	5	4	5	8	5	**27**	**57**
	TEST	6	6	6	6	6	**30**	
DN169	**TRAIN**	6	6	6	8	6	**32**	**57**
	TEST	5	5	5	5	5	**25**	
DN201	**TRAIN**	6	6	6	8	6	**32**	**67**
	TEST	7	7	7	7	7	**35**	
DN264	**TRAIN**	8	8	8	8	8	**40**	**80**
	TEST	8	8	8	8	8	**40**	
EN-B0	**TRAIN**	7	7	7	7	7	**35**	**50**
	TEST	3	3	3	3	3	**15**	
EN-B1	**TRAIN**	3	2	3	4	3	**15**	**25**
	TEST	2	2	2	2	2	**10**	
EN-B2	**TRAIN**	4	3	4	5	4	**20**	**40**
	TEST	4	4	4	4	4	**20**	
EN-B3	**TRAIN**	6	5	6	6	6	**29**	**34**
	TEST	1	1	1	1	1	**5**	

Table 3-5 Ranking of models for the AD vs. CN task [101]

Classifier	Evaluation Phase	ACC Rank	Precision Rank	Recall Rank	AUC Rank	F-1 Score Rank	Total Rank	Grand Total Rank
DN121	**TRAIN**	4	2	4	6	4	**20**	**51**
	TEST	7	3	7	7	7	**31**	
DN169	**TRAIN**	7	7	7	7	7	**35**	**50**
	TEST	3	4	2	3	3	**15**	
DN201	**TRAIN**	5	5	5	5	5	**25**	**64**
	TEST	8	7	8	8	8	**39**	
DN264	**TRAIN**	8	8	8	8	8	**40**	**66**
	TEST	5	6	5	5	5	**26**	
EN-B0	**TRAIN**	6	6	6	2	6	**26**	**55**
	TEST	6	5	6	6	6	**29**	
EN-B1	**TRAIN**	3	4	3	3	3	**16**	**40**
	TEST	4	8	4	4	4	**24**	
EN-B2	**TRAIN**	2	3	2	4	2	**13**	**19**
	TEST	1	2	1	1	1	**6**	
EN-B3	**TRAIN**	2	1	1	1	1	**6**	**15**
	TEST	**2**	**1**	**2**	**2**	**2**	**9**	

Table 3-6 Ranking of models for the AD vs. sMCI task [101]

For AD vs. CN, DenseNet264 had the highest score in both training and testing. DenseNet264 outperformed other models in AD vs. sMCI training, however DenseNet201, 121 and EfficientNet-B0 outscored DenseNet264 in testing. Overall, DenseNet264 had a superior position. DenseNet201, DenseNet121, and EfficientNet-B0 might be used to experiment with additional training data to develop a generalizable DL model for AD vs. sMCI task . EfficientNet-B2 and B3 were the worst performers in both categories.

3.4.2 Comprehensive Indicators and Efficiency-Effects Graph

We applied **Yang et al.'s method [46]** to examine the models' strengths in a more exhaustive manner in order to quantify the performance of the eight models with more precision. Since the F1-score is a combined measure of precision and recall, it is not included in this procedure. The dispersion and standard deviations (std) of the four indicators (precision, recall, accuracy, and

AUC) were determined. Each model's four indicators were added (sum),then determined their standard deviation (std). Finally, a constant (k = 0.04) was added to the standard deviation to prevent a division by zero error while calculating the comprehensive indicators. To get the values of comprehensive indicator, we divided the last two integers (sum/ (std + 0.04)). This method is shown in **Tables 3-7 and 3-8**.

.DL Model	Evaluation Phase	ACC	Precision	Recall	AUC	sum	std	std + .04	Comprehensive Indicator
DN121	**TRAIN**	0.99	0.99	0.99	1.00	3.98	0.00	0.04	**91.02**
	TEST	0.97	0.97	0.97	0.97	3.89	0.00	0.04	**96.39**
DN169	**TRAIN**	1.00	1.00	1.00	1.00	3.99	0.00	0.04	**93.84**
	TEST	0.93	0.94	0.93	0.93	3.73	0.01	0.05	**78.92**
DN201	**TRAIN**	1.00	1.00	1.00	1.00	3.99	0.00	0.04	**93.84**
	TEST	0.98	0.98	0.98	0.98	3.93	0.00	0.04	**97.27**
DN264	**TRAIN**	1.00	1.00	1.00	1.00	4.00	0.00	0.04	**100.00**
	TEST	1.00	1.00	1.00	1.00	3.98	0.00	0.04	**99.43**
EN-B0	**TRAIN**	1.00	1.00	1.00	1.00	3.99	0.00	0.04	**97.48**
	TEST	0.91	0.93	0.91	0.91	3.66	0.01	0.05	**73.20**
EN-B1	**TRAIN**	0.99	0.99	0.99	0.98	3.95	0.00	0.04	**93.72**
	TEST	0.87	0.90	0.87	0.87	3.51	0.02	0.06	**63.40**
EN-B2	**TRAIN**	0.99	0.99	0.99	1.00	3.97	0.00	0.04	**93.12**
	TEST	0.92	0.93	0.92	0.92	3.68	0.01	0.05	**75.48**
EN-B3	**TRAIN**	1.00	1.00	1.00	1.00	3.98	0.00	0.04	**96.70**
	TEST	0.86	0.89	0.86	0.86	3.46	0.02	0.06	**61.30**

Table 3-7 Comprehensive Indicators of Model Performance Measures for AD vs. CN [101]

In addition to the comprehensive indicators, the number of model parameters was used to evaluate the models' strengths. Figures 3-9 to 3-12 depict the efficiency-effects plot, where the vertical coordinate is the model's comprehensive indicator, and the horizontal coordinate is the number of model parameters. If the point that represents the model is close to the top-left corner, the model will be better and more effective. The opposite result is produced by the models in the bottom-right corner.

DL Model	Evaluation Phase	ACC	Precision	Recall	AUC	sum	std	std + .04	Comprehensive Indicator
DN121	**TRAIN**	0.79	0.79	0.78	0.85	3.21	0.03	0.07	**43.49**
	TEST	0.82	0.83	0.82	0.82	3.28	0.00	0.04	**72.89**
DN169	**TRAIN**	0.81	0.83	0.81	0.86	3.32	0.02	0.06	**52.23**
	TEST	0.80	0.83	0.80	0.80	3.22	0.02	0.06	**57.56**
DN201	**TRAIN**	0.79	0.79	0.79	0.83	3.20	0.02	0.06	**52.19**
	TEST	0.82	0.84	0.82	0.82	3.30	0.01	0.05	**68.44**
DN264	**TRAIN**	0.83	0.84	0.83	0.88	3.37	0.02	0.06	**52.46**
	TEST	0.81	0.83	0.81	0.81	3.26	0.01	0.05	**63.62**
EN-B0	**TRAIN**	0.81	0.81	0.81	0.81	3.24	0.01	0.05	**71.55**
	TEST	0.81	0.83	0.81	0.81	3.27	0.01	0.05	**68.64**
EN-B1	**TRAIN**	0.78	0.79	0.78	0.83	3.18	0.02	0.06	**52.63**
	TEST	0.81	0.84	0.81	0.81	3.27	0.02	0.06	**55.35**
EN-B2	**TRAIN**	0.78	0.79	0.78	0.83	3.18	0.03	0.07	**48.60**
	TEST	0.74	0.80	0.74	0.74	3.01	0.03	0.07	**43.45**
EN-B3	**TRAIN**	0.72	0.74	0.72	0.73	2.91	0.01	0.05	**58.00**
	TEST	0.75	0.79	0.75	0.75	3.04	0.02	0.06	**49.03**

Table 3-8 Comprehensive Indicators of Model Performance Measures for AD vs. sMCI [101]

The DenseNet-121 model showed the best testing performance for both kinds of categorization, while the EfficientNet-B0 model had the best overall metrics and lowered model parameters. When comparing evaluation matrices for AD vs. CN classification, DenseNet264 performed better than the competition. However, since it had the most parameters, it was more difficult to train. The DenseNet169 and EfficientNetB1, B2, and B3 models faired modestly in terms of overall metrics. DenseNet201 performed exceptionally well in testing, coming close to DenseNet121 and outperforming DenseNet264 for both kinds of categorization.

It may be concluded that increased model performance is not necessarily associated with a larger number of model parameters.

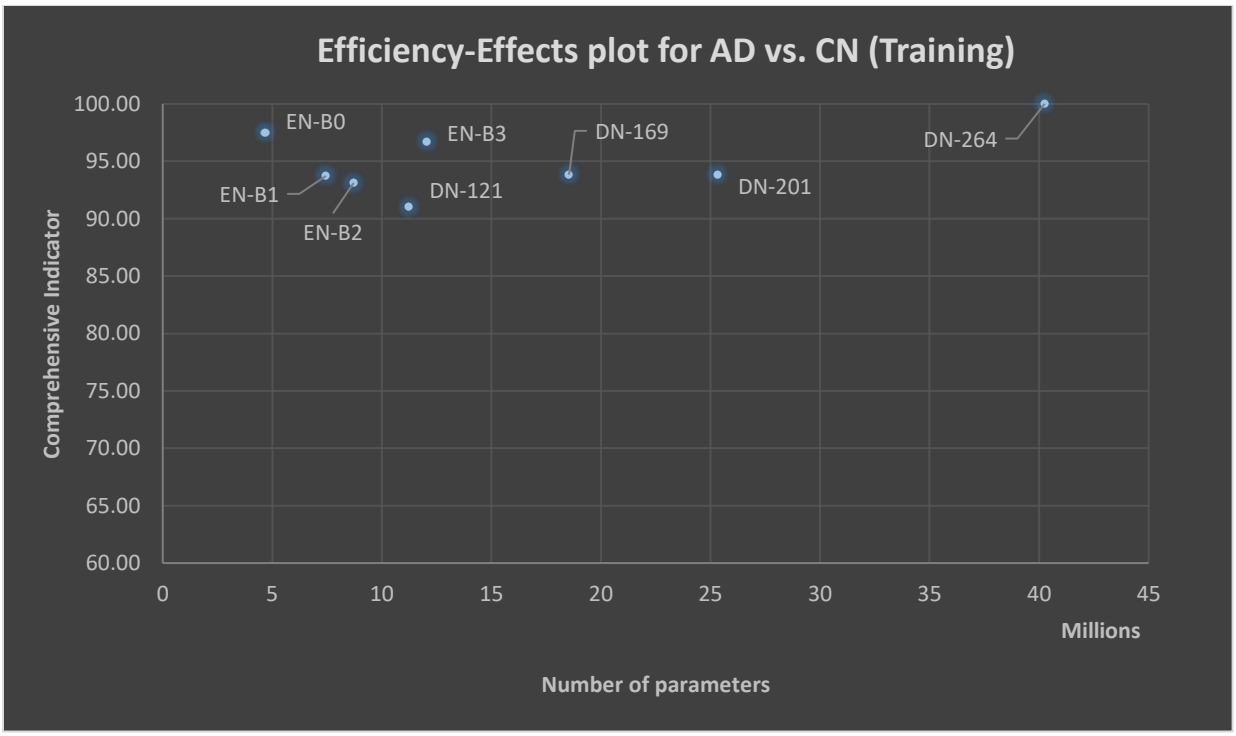


Figure 3-9 Efficiency Effects Graph for AD vs. CN (Training)

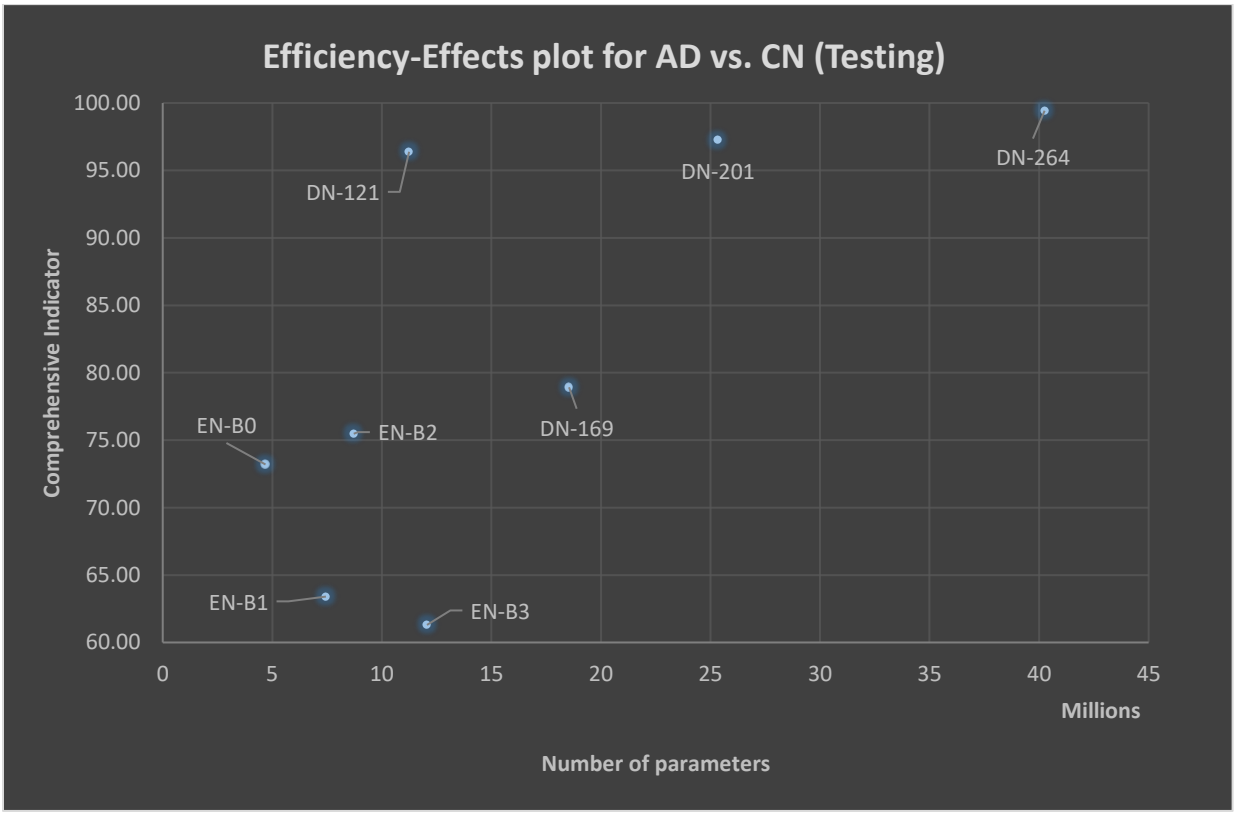


Figure 3-10 Efficiency Effects Graph for AD vs. CN (Testing)

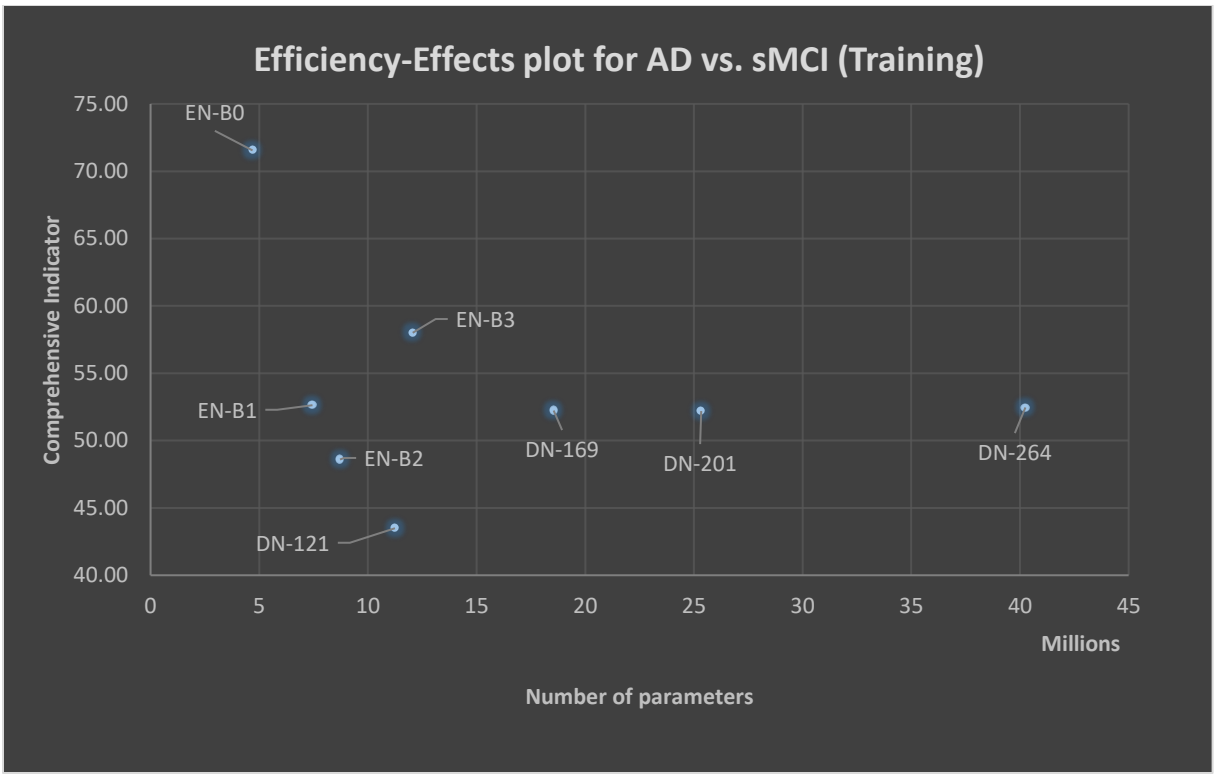


Figure 3-11 Efficiency Effects Graph for AD vs. sMCI (Training)

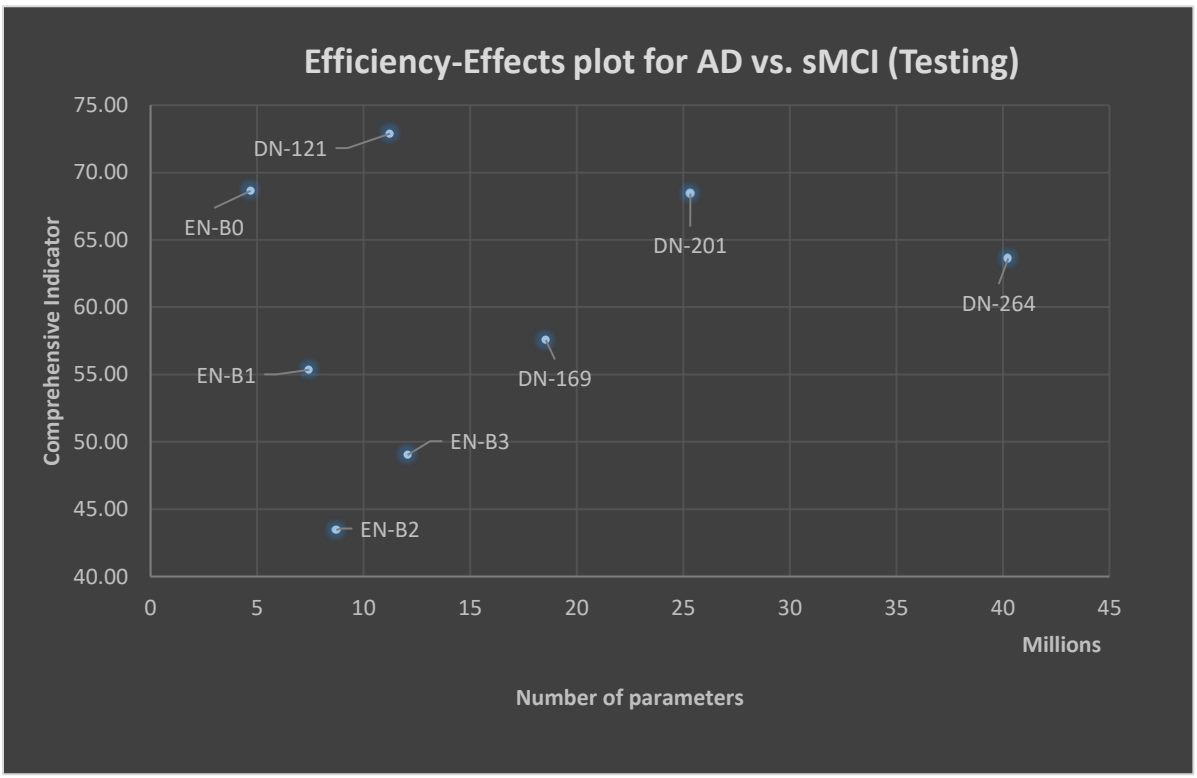


Figure 3-12 Efficiency Effects Graph for AD vs. sMCI (Testing)

3.5 Comparison with the Current state-of-the art methods

We compared our categorization findings to those in the literature, as shown in **Tables 3-9 and 3-10**. The methods evaluated varied from learning Level 1 to learning Level 3. We also compared our results for AD vs. sMCI to those for non-converter (stable) MCI vs. converter MCI, since converted MCI represents AD and non-converter MCI represents stable MCI.

We were able to discriminate AD from CN with the maximum degree of accuracy using E2EL, and consequently, this classifier may be employed in clinical settings after some qualitative study. The distinction between AD and sMCI, on the other hand, is more difficult than that between AD and CN because the morphological changes that must be detected are more subtle [50], [98]. The accuracy of several of the study results in Table 3–10 ranged from 70 to 80%. Our model performed well in this category, ranking first among Level 3 learning classifiers. The accuracy of AD vs. sMCI classification must be enhanced further to help in clinical contexts. This may be accomplished by employing additional training data, using a pre-trained 3D model, or using local TL.

3.6 Conclusion

We did experiments in this chapter to achieve **#Goal 2**, as illustrated in **Figure 1-1**. Even with the endemic issues of neuroimaging, where training data are few and sample dimensionality is high, DL models with E2EL can be used to obtain good accuracy. The experimental findings on the ADNI and IXI data showed that our model outperformed current state-of-the-art models in terms of performance and efficiency.

The findings of this chapter may be used in determining the best model to use and understanding the situations in which the models would produce better outcomes. In general, neural networks from the DenseNet family, such as DenseNet 121, DenseNet 201, DenseNet 264, and EfficientNet-B0, provide superior results for categorizing the various phases of AD.

EfficienNet-B0 because performed best overall for the AD vs. sMCI task, as can be seen in the Efficiency Effect Graph in Figure 3-11and Figure 3-12; it has fewer parameters; and its architecture has been developed to enable TL and make classifiers more generalizable.

Based on the findings of this chapter, we apply EfficientNet-B0 and a unique fusion of E2EL and TL in Chapter 4 to attempt to increase the accuracy of the AD vs. sMCI challenge.

Ref.	Learning Level/ Classifier	Subjects	AD vs. CN					
			Accuracy	Precision	SEN	AUC	F-1	SPE
Toshkhujaev et al.[185]	L1/RBF-SVM	AD-71, CN-171	91.57	-	81.82	-	-	100
Suk et al.[85]	L1/Regression+ CNN	AD-186,CN-286	91.02 ± 4.29	-	92.72	92.72	-	89.94
Zhang et al.[15]	L2/CNN	AD-280,CN-275	97.35	-	97.10	99.70	-	97.95
Li et al.[138]	L2/ CNN+RNN	AD-194,CN-216	89.10	-	84.6	91.0	-	93.1
Mehmood et al. [190]	L2/VGG-19(2D TL)	AD-85,CN-75	98.73	-	98.19	-	-	99.09
Tuan et al.[191]	L2/CNN+SVM	CN-98,AD-99	89.00	-	-	-	-	-
Song et al.[192]	L2/3D CNN	CN-126,AD-95	94.11	-	-	-	-	-
Nanni et al.[198]	L2/AlexNet^{P}	AD-137,CN-162	-	-	-	90.8	-	-
	L2/GoogleNet^{P}		-	-	-	89.6	-	-
	L2/ResNet50^{P}		-	-	-	89.8	-	-
	L2/ResNet101^{P}		-	-	-	89.9	-	-
	L2/InceptionV3^{P}		-	-	-	88.8	-	-
	L2/3DCNN		-	-	-	84.1	-	-
A et al.[20]	L2/2D CNN	CN-635,AD-220	96.8	-	94.0	-	-	96.0
Li et al.[169]	L2/CNN	CN-216,AD-194	85.9	-	81.5	88.4	-	89.9
Cui and Liu.[170]	L2/3DCNN	CN-223,AD-192	92.29	-	90.63	96.95	-	93.72
Liu et al.[204]	L2/2DCNN	CN-100,AD-93	93.26	-	92.55	95.68	-	93.94
Xu et al.[205]	L2/SRC	CN-117,AD-113	94.8	-	95.6		-	94.0
Pan et al.[206]	L2/CNN	AD-237,CN-242	93.75	-	91.49	96.87	-	95.92
Shi et al.[209]	L2/MM-SDPN	AD-51, CN-52	97.13± 4.44	-	95.93±7.84	-	-	95.93±7.84
Lu e al.[144]	L2/MDNN & TL	CN-304,AD-226	93.58	-	91.54	-	-	95.06
Hon and Khan[147]	L2/InceptionV4	AD-200,CN-100	96.25	-	-	-	-	-
Liu et al.[133]	L3/3D CNN	AD-97, CN-119	88.9	-	86.6	92.5	-	90.8
Oh et al. [141]	L3/CAE+3DCNN	CN-230,AD-198	86.60 ± 3.66	-	88.55	-	-	84.54
Proposed	***L3/DenseNet264***	***CN-245,AD-245***	***99.55***	***99.56***	***99.55***	***99.55***	***99.55***	***99.55***

Table 3-9 Performance of Published State-of-the-Art Methods for AD vs. CN Task [101]

P: Pertained, MM-SDPN: multimodal stacked deep polynomial networks,
MDNN: Multistate Deep Neural Network, SRC: sparse representation-based classification,
MiSePyNet: Multi-view Separable Pyramid Network.

Ref.	Learning Level/ Classifier	Subjects	AD vs. stable MCI					
			Accuracy	Precision	Recall	AUC	F-1	SPE
Suk et al.[85]	L1/ Regression+ CNN	pMCI-167,sMCI-226	74.82 ± 6.80	-	**70.93**	75.39	-	**78.82**
Zhang et al.[15]	L2/CNN	pMCI-162,sMCI-251	78.79	-	75.16	86.79	-	82.42
Li et al.[138]	L2/ CNN+RNN	pMCI-164,sMCI-233	72.5	-	61.0	74.6	-	82.5
Nanni et al.[198]	L2/AlexNet^{P}	sMCI-234,pMCI-240	-	-	-	69.1 ± 1.3	-	-
	L2/GoogleNet^{P}		-	-	-	70.0 ± 1.3	-	-
	L2/ResNet50^{P}		-	-	-	70.4 ± 1.0	-	-
	L2/ResNet101^{P}		-	-	-	71.2 ± 1.2	-	-
	L2/InceptionV3^{P}		-	-	-	69.8 ± 3.5	-	-
	L2/3DCNN		-	-	-	61.1 ± 1.0	-	-
Li et al.[169]	L2/CNN	pMCI-164,sMCI-233	71.0	-	59.8	71.9	-	79.0
Cui and Liu.[170]	L2/3DCNN	sMCI-231,pMCI-	75.00	-	73.33	77.70	-	76.19
Xu et al.[205]	L2/SRC	MCI-110	77.8	-	74.10		-	81.50
Pan et al.[206]	L2/MiSePyNet	sMCI-360,pMCI-166	**83.81**	-	**75.76**	**88.89**	-	**87.50**
Shi et al.[209]	L2/MM-SDPN	pMCI-43,sMCI-56	78.88±4.38	-	68.04±9.99	-	-	86.81± 9.12
Lu e al.[144]	L2/MDNN & TL	sMCI-409,pMCI-112	**81.55**	-	**73.33**	-	-	**83.83**
Shen et al.[210]	L2/RNN	pMCI-307,sMCI-558	**80.00**	-	**81.00**	-	-	**80.00**
Yang and Liu[146]	L2/SVM	sMCI-270,pMCI-70	78.56	-	91.02	-	-	77.63
Gao et al.[139]	L3/3DCNN	pMCI-168,sMCI-129	76.0	-	77.0	81.0	-	76.0
Oh et al. [141]	L3/CAE+3DCNN	sMCI-101,pMCI-166	73.95 ± 4.82	-	77.46	-	-	70.71
Proposed	***L3/DenseNet264***	***sMCI-229,AD-229***	***82.50***	***84.10***	***82.50***	***87.63***	***82.15***	***82.50***

Table 3-10 Performance of Published State-of-the-Art Methods for AD vs. sMCI Task [101]

Chapter 4: Implementation of EfficientNet-B0 based on the Fusion of TL and E2EL

Summary : As shown in Figure 1-1, this chapter is about the **second contribution**, "Identification and implementation of the optimum DL model," intended to help achieve the **second goal**, "to provide the neuroradiologist with the computer-interpretable information they need to analyze an MRI scan."

DenseNet264 gave ample accuracy for the AD vs. CN task. However, the accuracy obtained for the AD vs. sMCI task was only 82.50%. To increase the accuracy for this task, we chose the EfficienNet-B0 model based on the recommendations of chapter 3.

We implemented the EfficietNet-B0 using a novel approach—"**fusion of E2EL and TL**"—for the **AD vs. sMCI task and obtained 95.29% accuracy and 95.35% AUC for training and 93.10% accuracy and 93.00% AUC for testing**.

Additionally, we have also implemented EfficientNet-B0 with E2EL **for multiclass AD vs. CN vs. sMCI classification task and obtained 85.66% accuracy and 86% AUC for model training and 87.38% accuracy and 88.00% AUC** for model testing. . In Chapter 5, **we will utilize this multiclass classifier to build an ensemble learning-based tool called "Deep-AD".**

4.1 Introduction

According to the literature, detecting AD at an early stage is crucial for patients to get its full benefits[41], [50]. Currently, 86.30% [140] is the highest accuracy level for this job whether using E2EL, local TL, CNN-based 2D transfer learning, or ROI segmentation techniques. Therefore, we suggest combining E2EL and TL during the model's training phase to improve accuracy as well as generalization capacity. By transferring the learning from each fold of the 5-fold stratified cross-validation to the fold after it, and so forth, we trained the EfficientNet-B0 CNN for the binary classification task of AD vs. sMCI classification . E2EL was used to train the model in the first fold. We also trained and evaluated an E2EL-based EfficientNet-b0 model for the multiclass AD vs. CN vs. sMCI classification problem. 3D T1W MRI scans were preprocessed and input into the models.

4.2 Materials and methods

We propose DL models for the classification of MRI scans of 1) AD and sMCI subjects, and 2) AD, sMCI, and CN subjects. For task one we used a novel E2EL and TL fusion approach, as shown in **Figure 4-1**. During model training via 5-fold stratified cross-validation, we trained the model from scratch in the first fold (E2EL), validated it, and used the final weights of the best epoch from fold 1 as the initial weights for fold 2 (TL). After training and validating the model in fold 2, we used the final weights of the best epoch from fold 2 as the initial weight for fold 3, and we repeated the same steps for the subsequent folds. For task 2, we used E2EL for all the folds. The model of the best epoch from each fold was used to assess external MRI scans to check for overfitting.

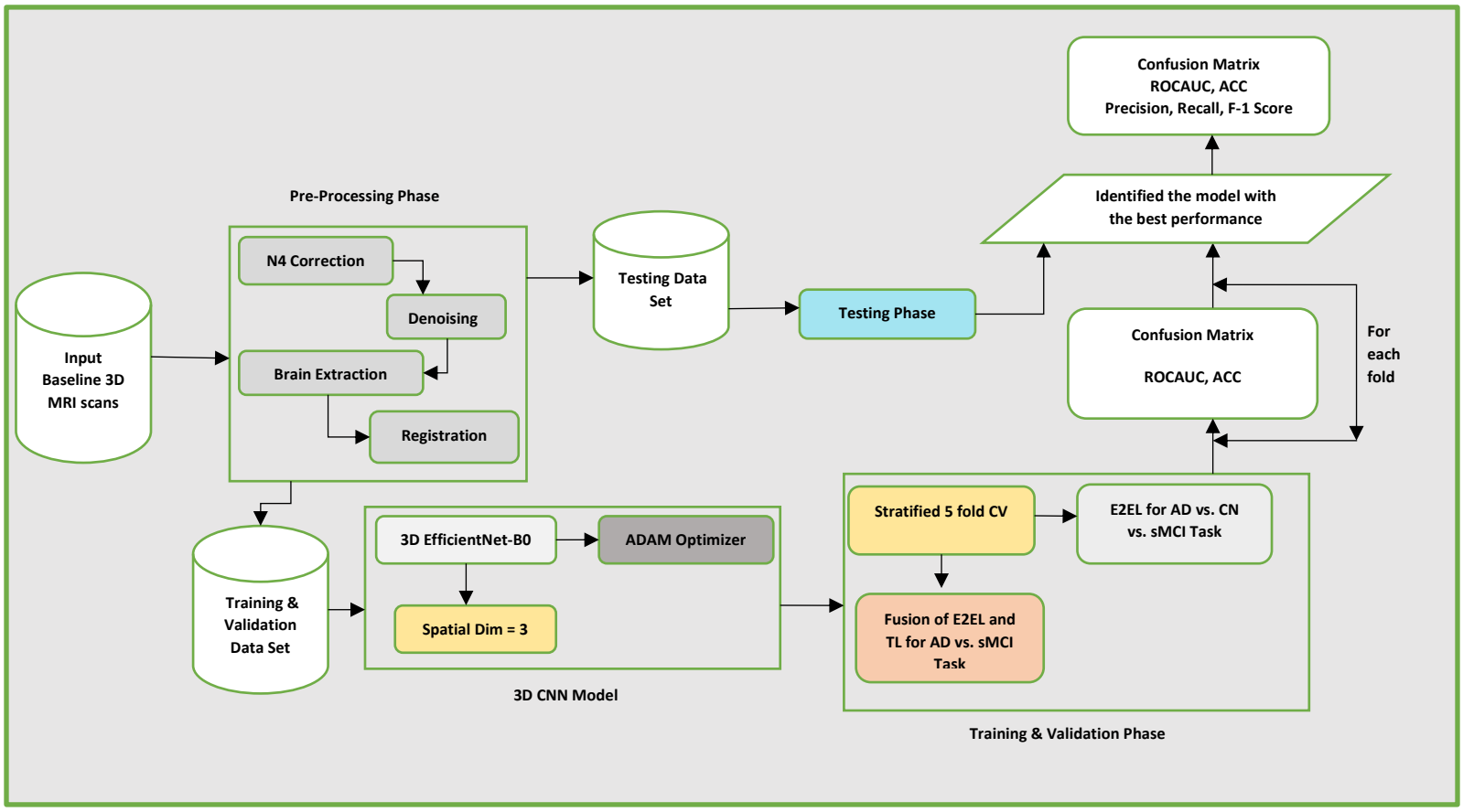


Figure 4-1 . Block diagram of proposed work

4.2.1 Participants

Same as discussed **3.2.1 Participants**.

4.2.2 Preprocessing Pipeline

Same as discussed in **3.2.2 Preprocessing Pipeline**

4.2.3 Implemented CNN

In recent years, CNNs have seen a surge in popularity because of their impressive usefulness in high-dimensional data analysis. EfficientNet models are based on simple and incredibly effective compound scaling methods. In many cases, EfficientNet models achieve better accuracy and efficiency than state-of-the-art CNNs like AlexNet, ImageNet, GoogleNet, or MobileNetV2 [24]. EfficientNets are more compact, run faster, and generalize more effectively, leading to improved accuracy. They have often been used with TL. However, they have only been pre-trained on 2D images, so their learning cannot be transferred to 3D MRI scans. Nonetheless, they can be trained for 3D scans via E2EL.

Models from b0 to b7 [220] are represented in EfficientNet, with individual parameter sets spanning from 4.6 to 66 million. We chose the EfficientNet-B0 model for the proposed classification tasks because it offered the best overall evaluation metrics and the lowest model parameters, as reported in chapter 3.

Figure 4-2 depicts the implemented EfficientNet-b0 structural layout. It has a total of 295 layers, distributed as shown in Table 4-1. Six consecutive blocks with various structures are included, in addition to 16 MBConvBlocks.

4.2.4 Exploratory Setup

For AD vs. sMCI and AD vs. CN vs. sMCI tasks, EfficientNet-B0 is implemented by using stratified 5-fold CV as shown in **Algorithm 3 and 4.**

4.2.4.1 Frameworks, Tools, and IDEs

Same as discussed in **3.2.4.1 Frameworks, Tools, and IDEs**

4.2.4.2 Network Hyperparameters

Same as discussed in **3.2.4.2 Network Hyperparameters**

4.2.4.3 Evaluation Matrices

Same as discussed in **3.2.4.3 Evaluation Matrices**

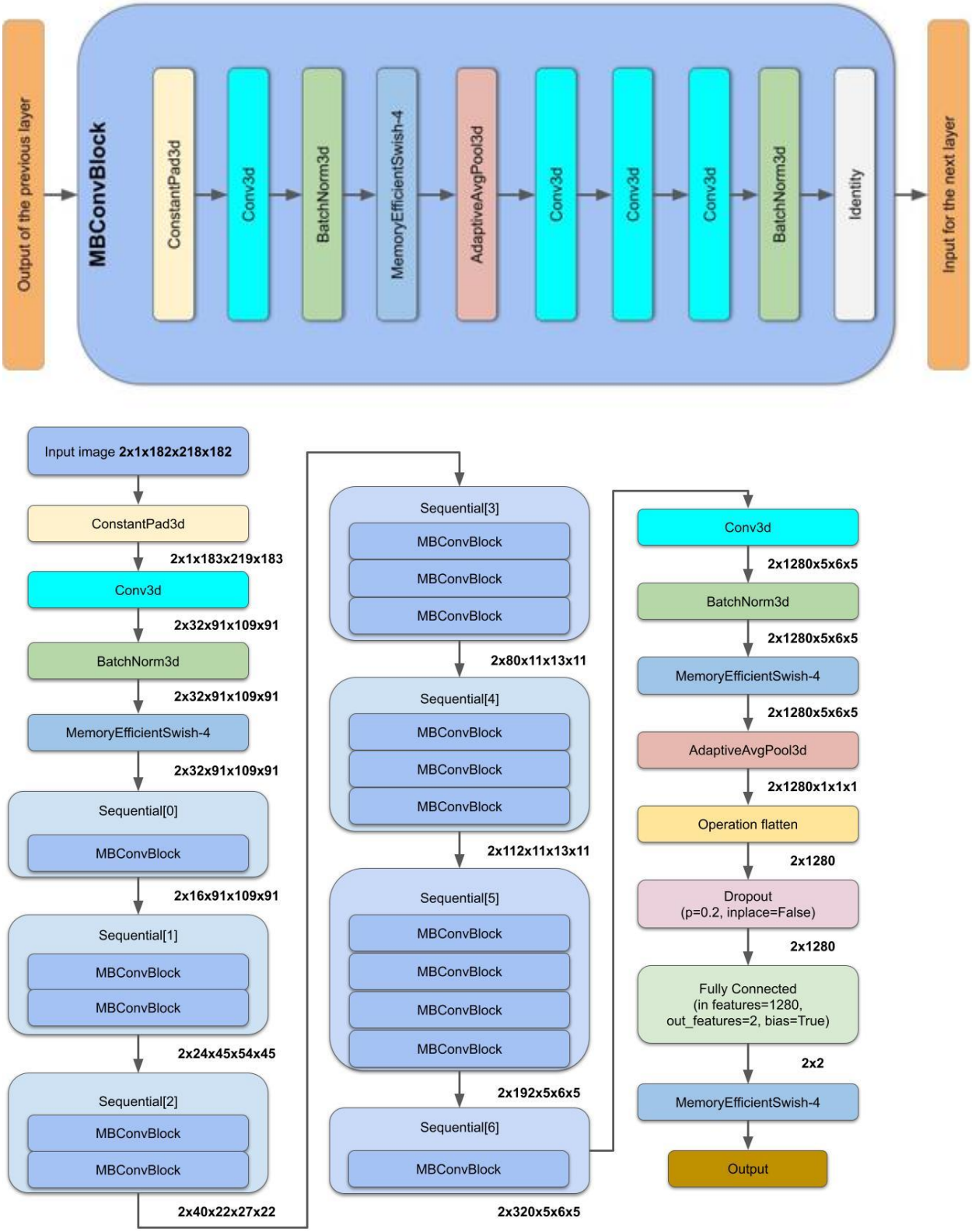


Figure 4-2 (a) Structural layout of EfficienNet-B0

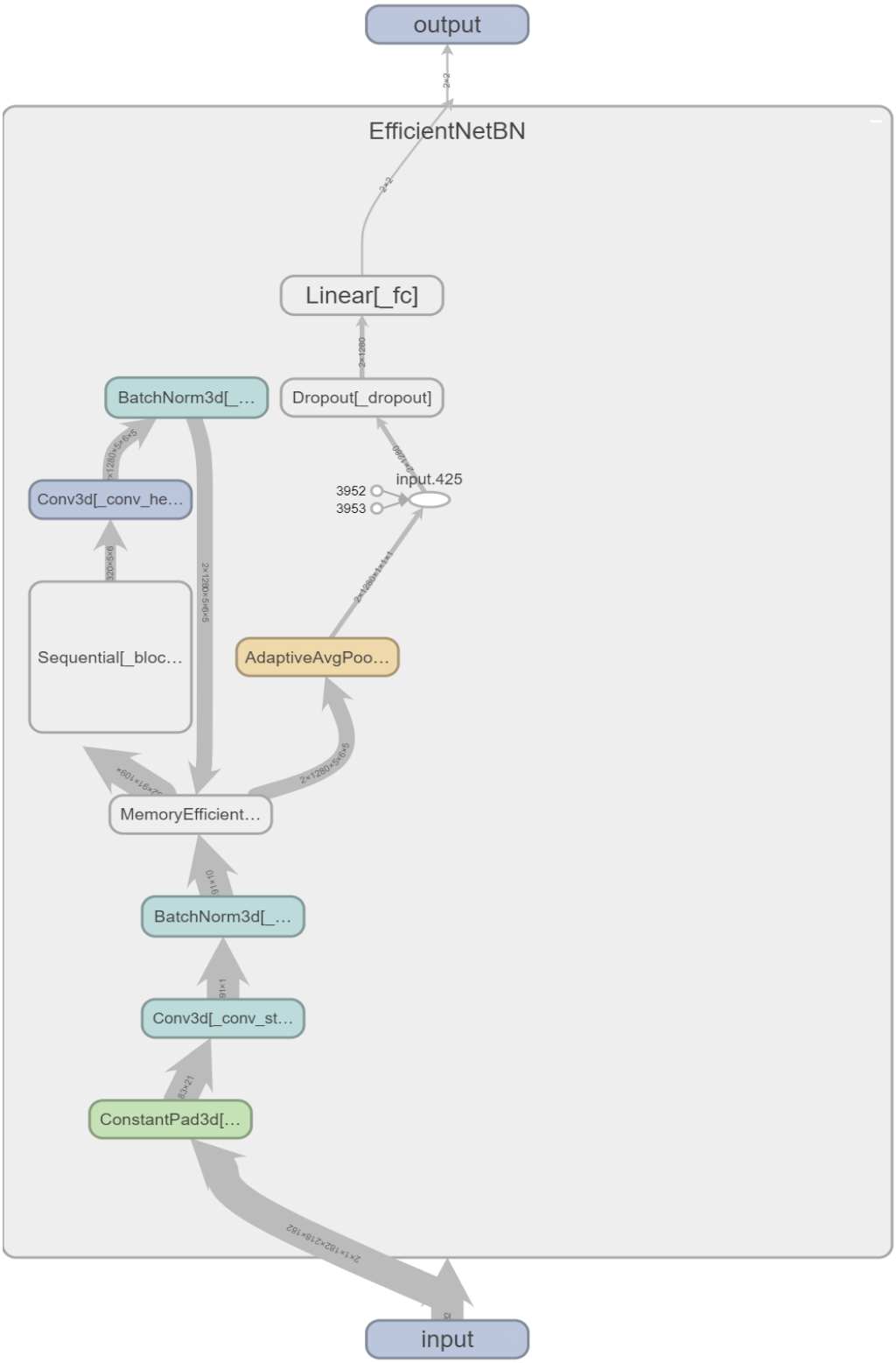


Figure 4-2 (b) Structural layout of EfficienNet-B0, drawn using the Tensor Board [80] in PyTorch

[80] https://pytorch.org/docs/stable/tensorboard.html

Input size	**[2,1,182,218,182]**
Layer (Type)	**Output Shape**
ConstantPad3d	[2, 1, 183, 219, 183]
Conv3d	[2, 32, 91, 109, 91]
BatchNorm3d	[2, 32, 91, 109, 91]
MemoryEfficientSwish-4	[2, 32, 91, 109, 91]
Sequential [0] block Contains one MBConvBlock **[ConstantPad3d,Conv3d, BatchNorm3d, MemoryEfficientSwish-4 ,AdaptiveAvgPool3d,again have three Conv3d and one BatchNorm3d,Identity]**	[2, 16, 91, 109, 91]
Sequential [1] block contains two MBConvBlocks.	[2,24,45,54,45]
Sequential [2] block contains two MBConvBlocks.	[2,40,22,27,22]
Sequential [3] block contains three MBConvBlocks.	[2,80,11,13,11]
Sequential [4] block contains three MBConvBlocks.	[2,112,11,13,11]
Sequential [5] block contains four MBConvBlocks.	[2,192,5,6,5]
Sequential [6] block contains one MBConvBlocks.	[2,320,5,6,5]
Conv3d	[2,1280,5,6,5]
BatchNorm3d	[2,1280,5,6,5]
MemoryEfficientSwish-4	[2,1280,5,6,5]
AdaptiveAvgPool3d	[2,1280,1,1,1]
Operation flatten	[2,1280]
Dropout(p=0.2, inplace=False)	[2,1280]
Fully Connected : Linear(in features=1280, out features=2, bias=True)	[2,2]
MemoryEfficientSwish()	Output
Total params: 4,690,942 **Trainable params: 4,690,942** **Non-trainable params: 0**	
Input size (MB): 27.55 **Forward/backward pass size (MB): 7754.59** **Params size (MB): 17.89** **Estimated Total Size (MB): 7800.03**	

Table 4-1 Summary of the 295 layers of the model with output shape

4.2.4.4 Algorithm

After the preprocessing of all data, the MRI scans were distributed as follows:

- For AD vs. sMCI task :
  - AD: 160 for training, 40 for validation, 29 for testing.
  - sMCI: 160 for training, 40 for validation, 29 for testing.
- For AD vs. CN vs. sMCI task :
  - AD: 160 for training, 40 for validation, 29 for testing.
  - CN: 160 for training, 40 for validation, 45 for testing.
  - sMCI: 160 for training, 40 for validation, 29 for testing.

The implemented method for the classification of AD vs. sMCI is presented in **Algorithm 3**. As indicated in steps 2 and 3, we utilized MRI scans from DATASET [1] for training and validation. In steps 48 and 49, we initialized the training and validation loaders from DATSET [C] MRI scans to be analyzed in fold C. In steps 43 to 46, we evaluated unseen MRI scans using the model with the best weights from each fold.

In the first fold model learned from scratch (E2EL), we did not reset the model in step 50; conversely, we raised the counter to reflect the fold's rise since we intended to apply this fold's learning to the next fold (TL), and so on for successive folds.

The same approach, with a few modifications, was applied to the AD, sMCI, and CN multiclassification task as shown in **Algorithm 4**: i.e., there were three types of input MRI scans; the batch size and worker values were set to eight; the number of output classes was changed to three; and the number of epochs was increased to 100. The same process was used for testing, validation, and training inside each fold. After each fold, the model was reset to apply E2EL.

Algorithm For sMCI vs. AD classification task

Require: Pre-processed T1 MP-RAGE MRI scans of AD and stable MCI
Ensure: A trained and validated 3D CNN model for improved inferencing .

batchsize ← 2,*workers* ← 2 ▷ **Parameters for Data Loaders**
loader ← *DataLoader*(*trainingdataset*[1])
validationloader ← *DataLoader*(*validationdataset*[1])
testingloader ← *DataLoader*(*testingdataset*)
device ← *cuda* ▷ *Parameters for the model*
spatial_dims ← 3 ▷ for 3D Input
input_channels ← 1 ▷ gray-scale images
num_classes ← 2 ▷ Number of output classes
optimizer ← *Adam*, *learning_rate* ← .0001 ▷ Tuned Hyper-parameter
loss_function ← **CrossEntropyLoss()**
auc_metric ← **ROCAUCMetric()**
model ← **EfficientNetBN("efficientnet-b0")**
NO_epoch ← 50,*C* ← 1 ▷ Tuned Hyper-parameter
while $C \leq 5$ **do** ▷ Outer loop for carrying out 5-fold stratified CV
  epoch ← 1 ,*best_metric* ← −1 ▷ *begin a PyTorch training*
  while *epoch* ≤ *NO_epoch* **do** ▷ Training for the 50 epoch
    epoch_loss ← 0
    model.train() ▷ sets the mode to train.
    for `batch_data in trainloader` **do** ▷ employ mini-batches
      inputs&labels ← *batch_data*
      optimizer.zero_grad() ▷ Gradients should be set to 0.
      outputs ← *model*(*inputs*) ▷ The model infers.
      loss ← *loss_function*(*outputs*, *labels*) ▷ Determine the loss.
      loss.backward() ▷ compute gradients during the backward pass.
      optimizer.step() ▷ update the learning weights.
      epoch_loss ← *epoch_loss*+*loss*
    end for
    model.eval() with torch.no_grad() ▷ sets the model to evaluate
    for `val_data in validationloader` **do** ▷ evaluation loop
      val_inputs&val_labels ← *val_data*
      val_pred ← *model*(*val_inputs*)
      val_loss ← *loss_function*(*val_pred*, *val_labels*)
    end for
    Determine the acc_metric() and auc_metric() values
    if *acc_metric* ≥best_metric **then** ▷ Save the best metric model.
      A ← *val_labels* , *B* ← *val_pred* , *best_metric* ← *acc_metric*
      torch.save(model.state_dict())
    end if
    epoch ←epoch+1
  end while ▷ Training Completed for FOLD C.
  draw_confusion_matrix(A,B),draw_roc(A,B)
  model.eval() with torch.no_grad() ▷ sets the model to test
  for `test_data in testingloader` **do** ▷ test using new MRI scans
    test_inputs&test_labels ← *test_data*
    test_pred ← *model*(*test_inputs*)
  end for
  draw_confusion_matrix(test_pred,test_labels),draw_roc(test_pred,test_labels)
  trainloader ← *DataLoader*(*trainingdataset*[$C+1$])
  validationloader ← *DataLoader*(*validationdataset*[$C+1$])
  C ←C+1 ▷ model is not reset; current fold learning will be applied in the following fold.
end while

Algorithm 3 Used for implementing EfficienNet-B0 for AD vs. sMCI tasks[fusion of E2EL & TL].

Algorithm 4 For AD vs. CN vs. sMCI classification task

Require: Pre-processed T1 MP-RAGE MRI scans of AD and sMCI and CN Subjects
Ensure: Trained and validated 3D CNN model for improved inference .

batchsize ← 8,*workers* ← 8 ▷ ***Parameters for Data Loaders***
trainloader ← *DataLoader*(*trainingdataset*[1])
validationloader ← *DataLoader*(*validationdataset*[1])
testingloader ← *DataLoader*(*testingdataset*)
device ← *cuda* ▷ *Parameters for the model*
spatial_dims ← 3 ▷ for 3D Input
input_channels ← 1 ▷ gray-scale images
num_classes ← 3 ▷ Number of output classes
optimizer ← *Adam*, *learning_rate* ← .0001 ▷ Tuned Hyper-parameter
loss_function ← **CrossEntropyLoss()**
auc_metric ← **ROCAUCMetric()**
model ← **EfficientNet-B0**
NO_epoch ← 100,*C* ← 1 ▷ Tuned Hyper-parameter
while *C* ≤ 5 **do** ▷ Outer loop for carrying out 5-fold stratified CV
    epoch ← 1 ,*best_metric* ← −1 ▷ *b*egin a PyTorch training
    while *epoch* ≤ *NO_epoch* **do** ▷ Training for the 100 epoch
        epoch_loss ← 0
        model.train() ▷ sets the mode to train.
        for `batch_data in trainloader` **do** ▷ employ mini-batches
            inputs&labels ← *batch_data*
            optimizer.zero_grad() ▷ Gradients should be set to 0.
            outputs ← *model*(*inputs*) ▷ The model infers.
            loss ← *loss_function*(*outputs*, *labels*) ▷ Determine the loss.
            loss.backward() ▷ compute gradients during the backward pass.
            optimizer.step() ▷ update the learning weights.
            epoch_loss ← *epoch_loss*+*loss*
        end for
        model.eval() with torch.no_grad() ▷ sets the model to evaluate
        for `val_data in validationloader` **do** ▷ evaluation loop
            val_inputs&val_labels ← *val_data*
            val_pred ← *model*(*val_inputs*)
            val_loss ← *loss_function*(*val_pred*, *val_labels*)
        end for
        Determine the acc_metric() and auc_metric() values
        if *acc_metric* ≥best_metric **then** ▷ Save the best metric model.
            A ← *val_labels* , *B* ← *val_pred* , *best_metric* ← *acc_metric*
            torch.save(model.state_dict())
        end if
        epoch ←epoch+1
    end while ▷ Training Completed for FOLD C.
    draw_confusion_matrix(A,B),draw_roc(A,B)
    model.eval() with torch.no_grad() ▷ sets the model to test
    for `test_data in testingloader` **do** ▷ test using new MRI scans
        test_inputs&test_labels ← *test_data*
        test_pred ← *model*(*test_inputs*)
    end for
    draw_confusion_matrix(test_pred,test_labels),draw_roc(test_pred,test_labels)
    trainloader ← *DataLoader*(*trainingdataset*[*C* + 1])
    validationloader ← *DataLoader*(*validationdataset*[*C* + 1])
    C ←C+1 ▷ reset the model; End to End Learning.
end while=0

Algorithm 4 Used for implementing EfficientNet-B0s for AD vs. CN vs. sMCI [E2EL].

4.3 Results

Figure 4-3 and 4-4 depicts the loss experienced during training as well as the changes in validation accuracy of fold 1 for the AD vs. sMCI task. Training loss was progressively reduced while validation accuracy peaked at 88.75% in the first fold, indicating that the model was learning adequately. This effect increased in succeeding folds because of the application of TL from prior folds, as can be seen **in Figure 4-5**.

Maximum testing accuracy and AUC reached 93.10% and 93.0% in fold 5, while average training accuracy over all folds reached 95.29%. The confusion matrix and ROCAUC for the optimal training and testing fold of the binary classification task are shown in **Figure 4-6 and 4-7,** respectively.

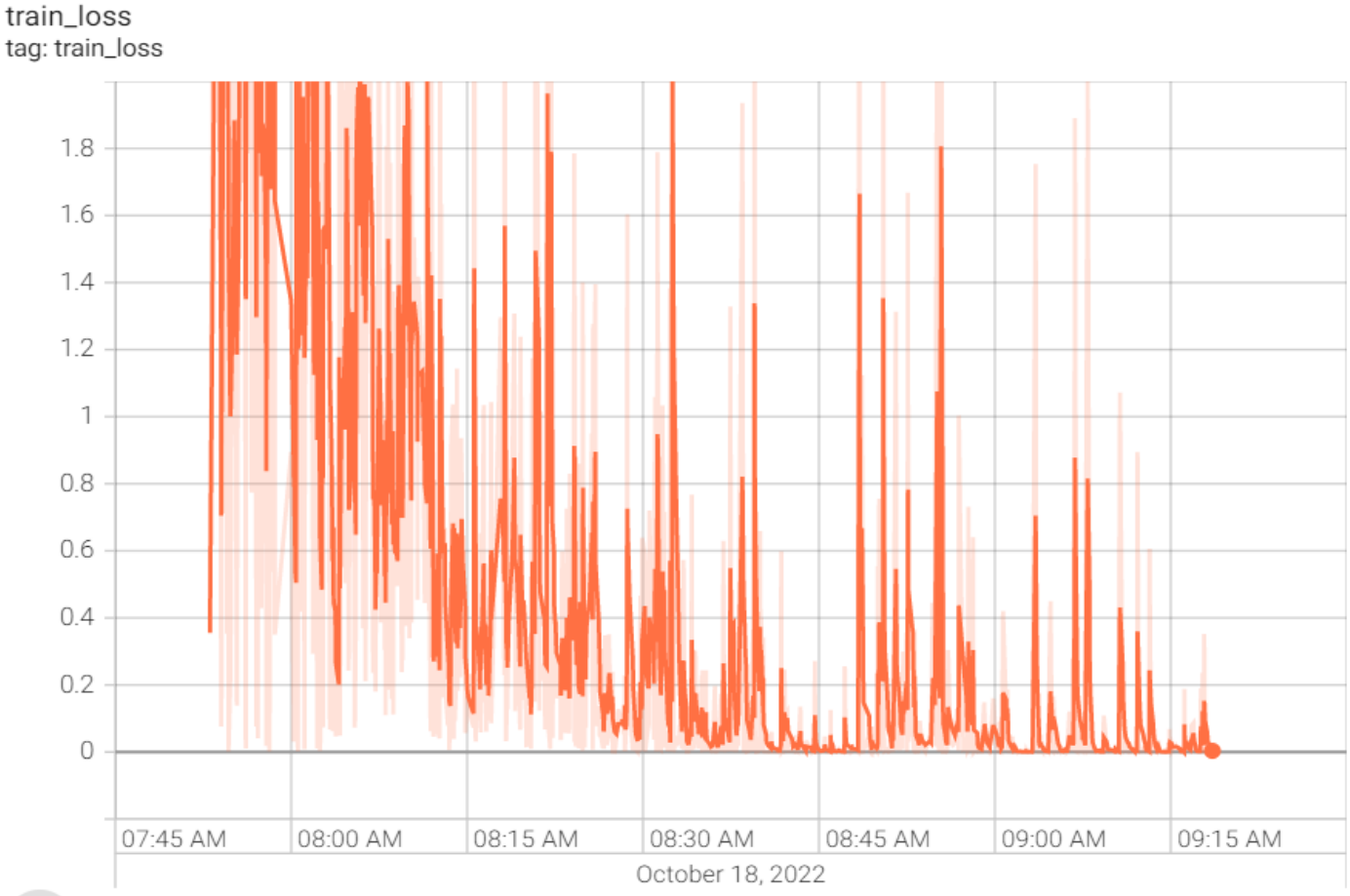


Figure 4-3 loss during the training in first fold for AD vs. sMCI task

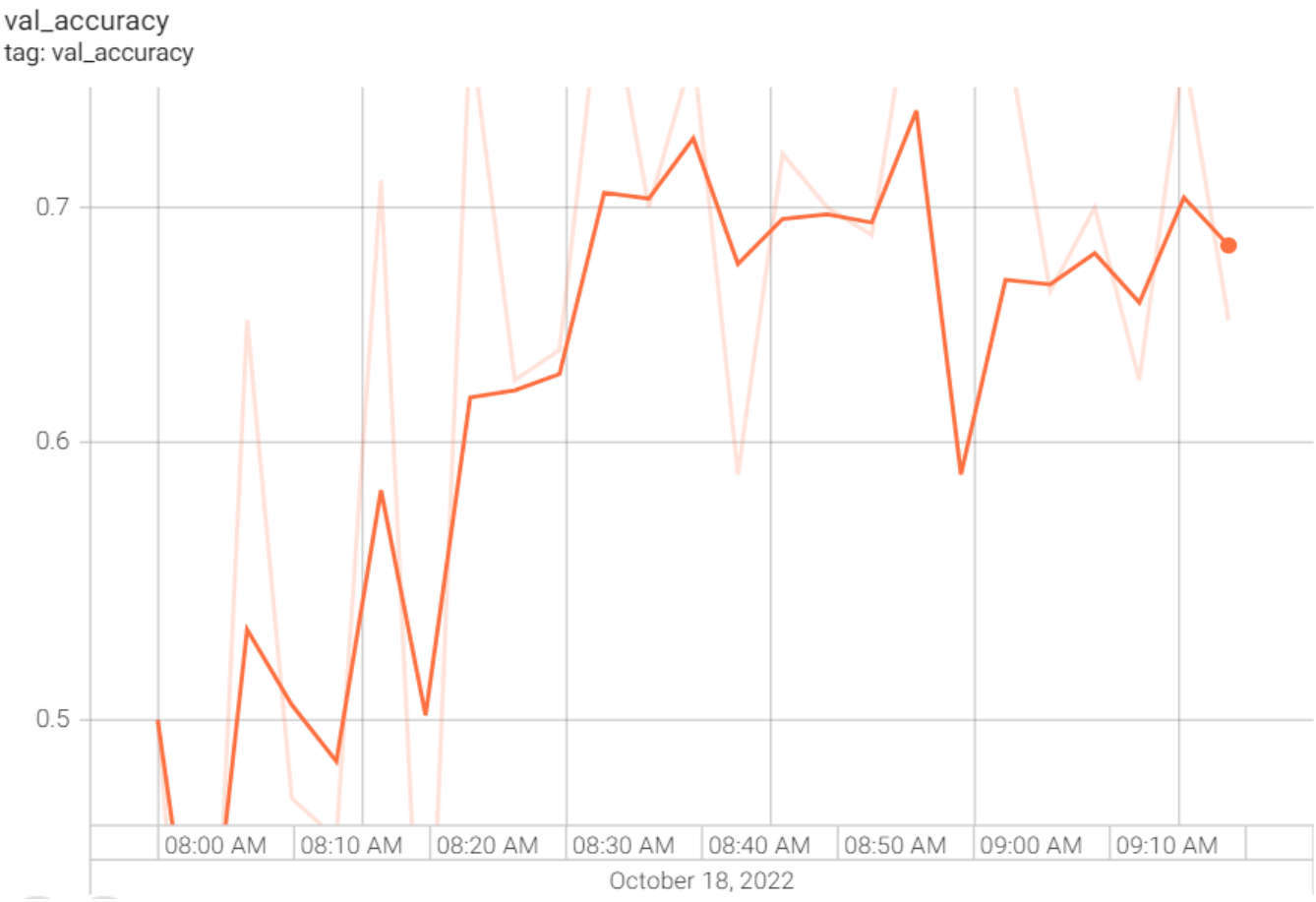


Figure 4-4 Validation accuracy in first fold for AD vs. sMCI task

Validation was performed on a total of 80 MRI scans, with 40 scans utilized for each class. As shown in **Figure 4-6,** TP for AD subjects was 40 and FN was 0, while TN for sMCI subjects was 36 and FP was 4. Due to the lack of information in the ADNI database regarding the status of each subject after x years of stability and as they present almost the same anatomical structures as AD scans, only a small number of sMCI subjects were misclassified. Both classes obtained an AUC of 95.0%.

Testing was performed on a total of 58 MRI scans (29 of AD and 29 of sMCI subjects) which had not been used during the training or validation procedures. As shown in **Figure 4-7**, TP for AD subjects was 29 and FN was 0, while TN for sMCI subjects was 25 and FP was 4. Both classes obtained an AUC of 93.0%.

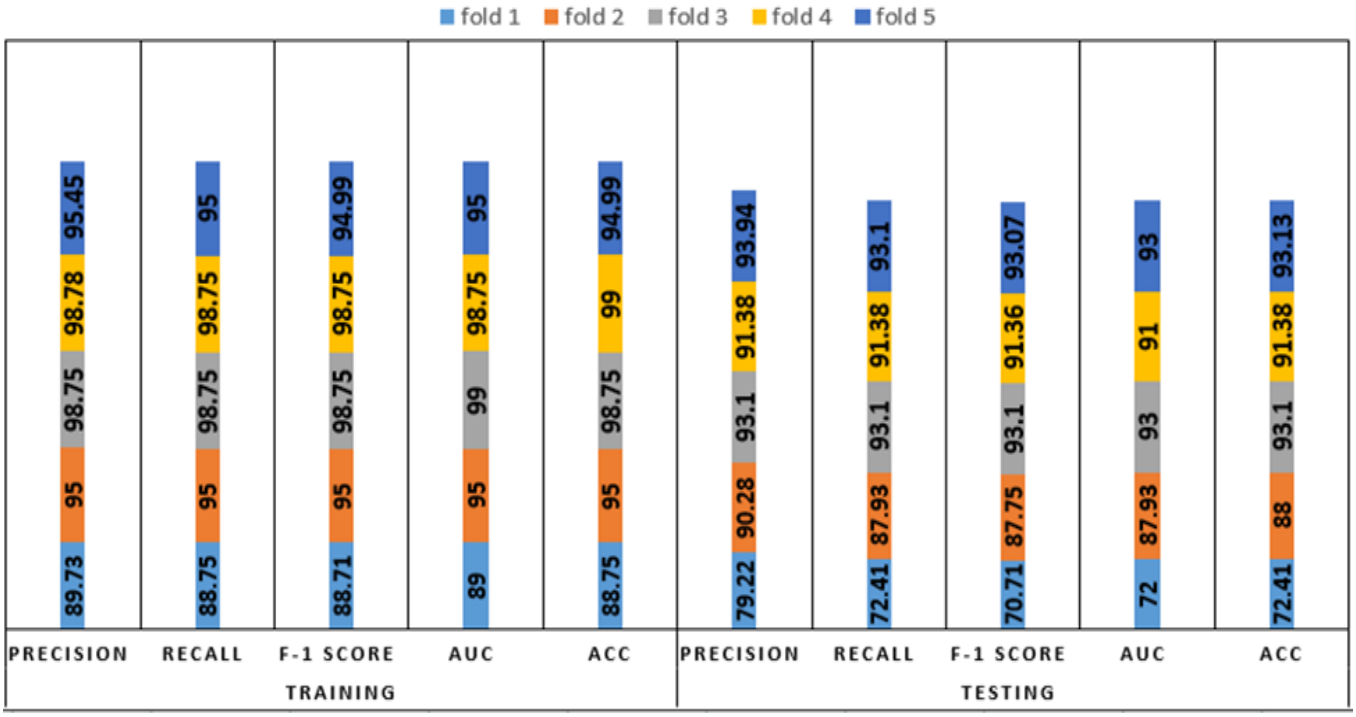


Figure 4-5 Comprehensive results for the AD vs. sMCI task

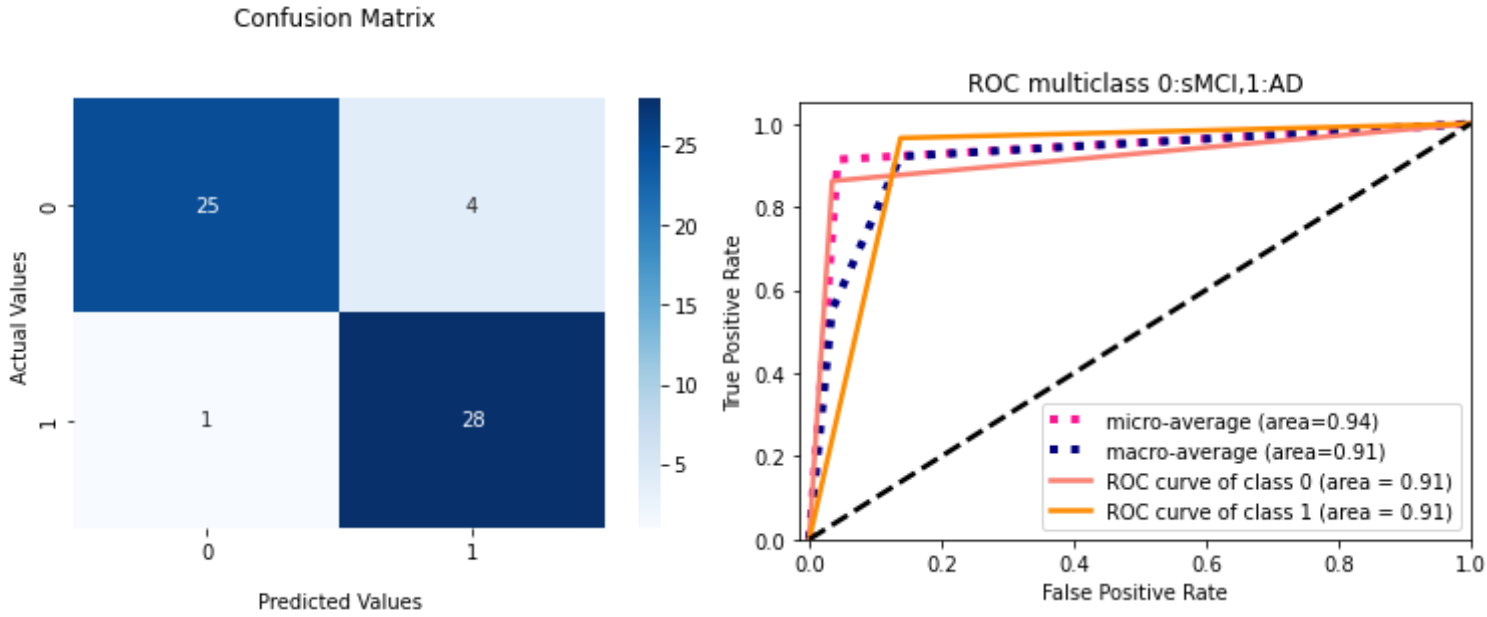


Figure 4-6 ROCAUC Curve/ Confusion Matrix of the optimal training fold for the AD vs. sMCI task.

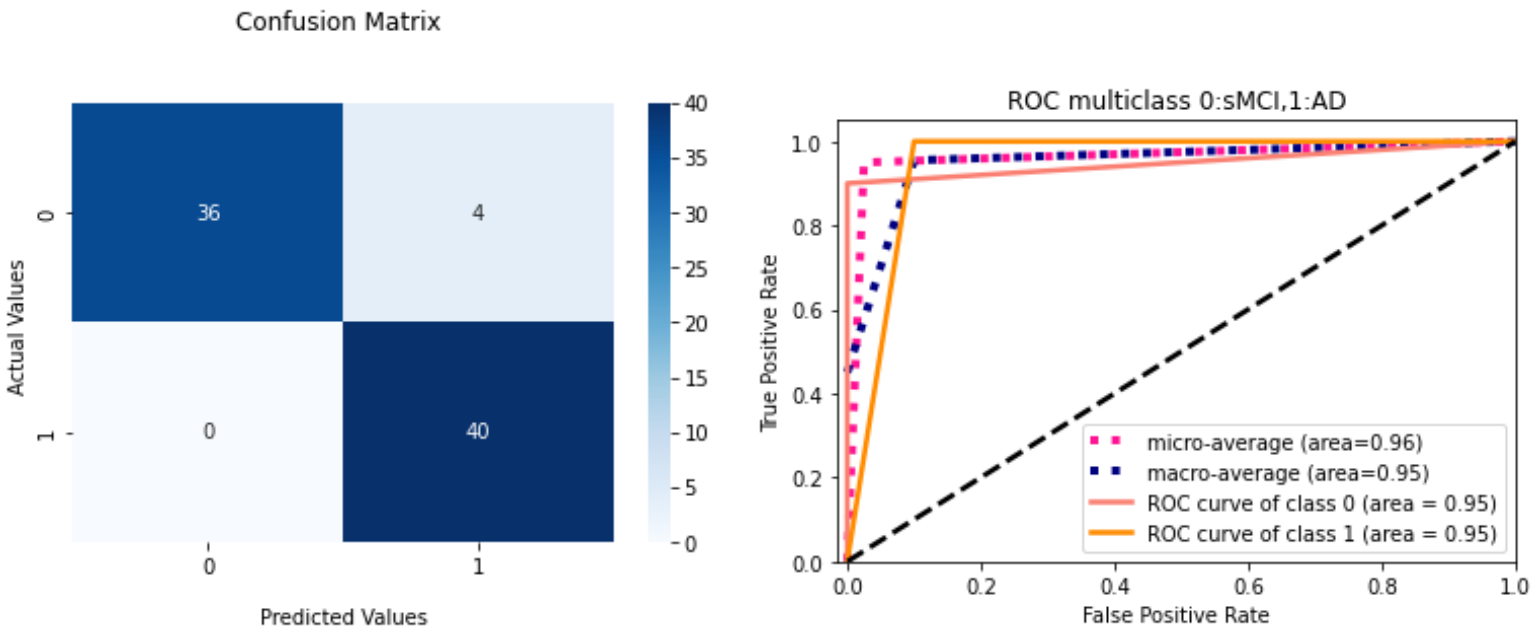


Figure 4-7 ROCAUC Curve/ Confusion Matrix of optimal testing fold for the AD vs. sMCI task.

Figure 4-8 and 4-9 depicts the training loss as well as variations in validation accuracy for the AD vs. sMCI vs. CN task. Training loss was progressively reduced, and validation accuracy peaked at 85.83% in the first fold, implying that the model was learning appropriately.

As shown in **Figure 4-10**, accuracy is altered in the subsequent folds due to learning from scratch in every fold. In fold 2, the highest testing accuracy reached 87.38%, while average training accuracy across all folds reached 85.66%.

The confusion matrix and ROCAUC for the optimal training and testing fold of the multiclassification task are shown in **Figures 4-11 and 4-12, respectively**. Validation was performed on a total of 120 MRI scans, with 40 scans utilized for each class, nearly 100% accuracy was achieved for CN subjects; however, because of their similar anatomical structures, the categorization findings for the AD and sMCI participants indicated a small number of incorrect classifications. The corresponding AUCs for the CN, AD, and sMCI classes were 100.0%, 88.0%, and 84%, respectively.

Testing was performed on a total of 103 MRI scans (29 of AD, 29 of sMCI, and 45 of CN subjects) which had not been used during training or validation. As shown in **Figure 4-12**, quite a few sMCI and CN participants were incorrectly classified, while AD subjects were 100% correctly classified. The corresponding AUCs for the CN, AD, and sMCI classes were 93.0%, 95.0%, and 84%, respectively.

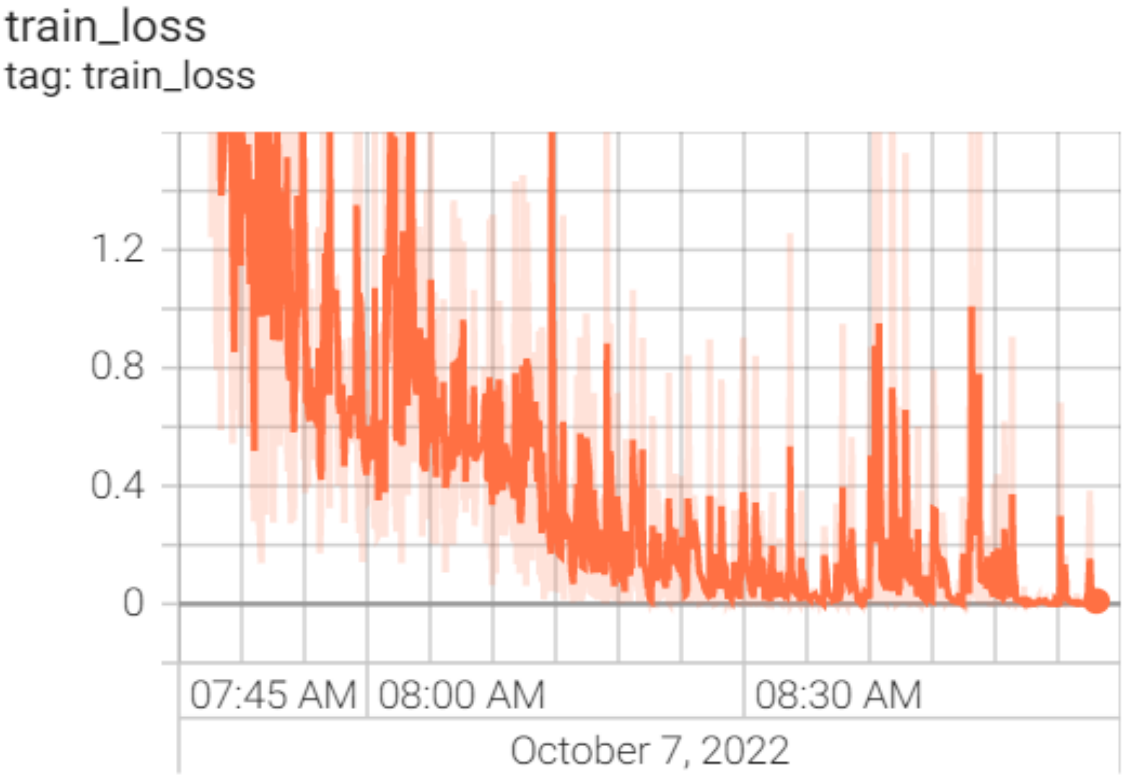


Figure 4-8 loss during the training in first fold for AD vs.CN vs. sMCI task

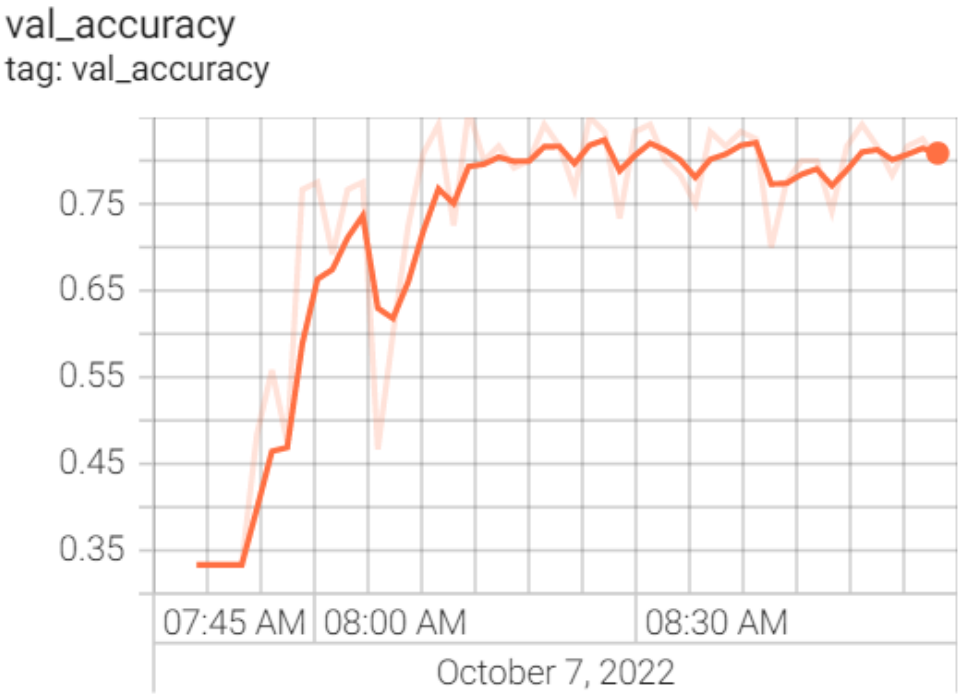


Figure 4-9 Validation accuracy in first fold for AD vs. CN vs. sMCI task

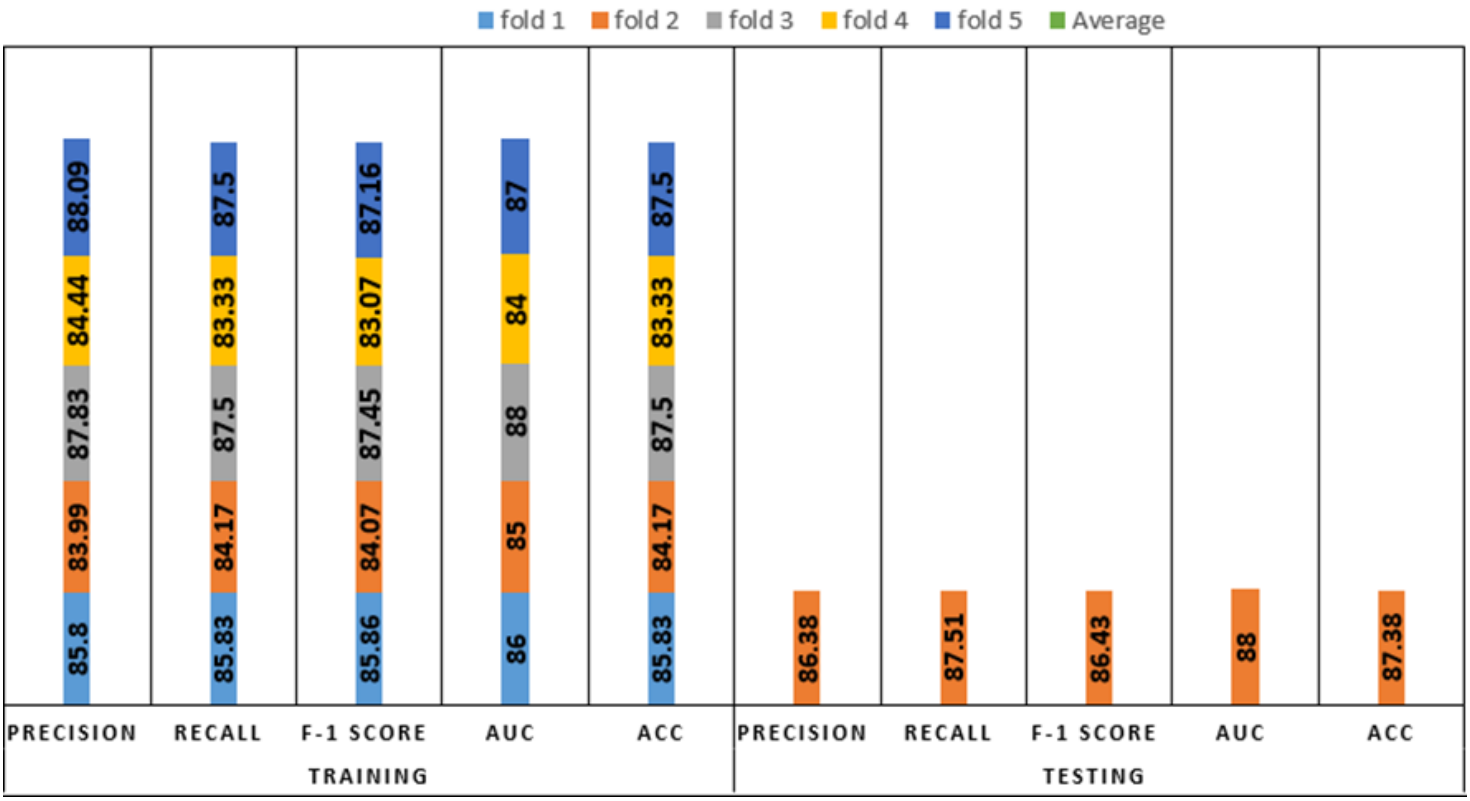


Figure 4-10 Comprehensive results for the AD vs. CN vs. sMCI task

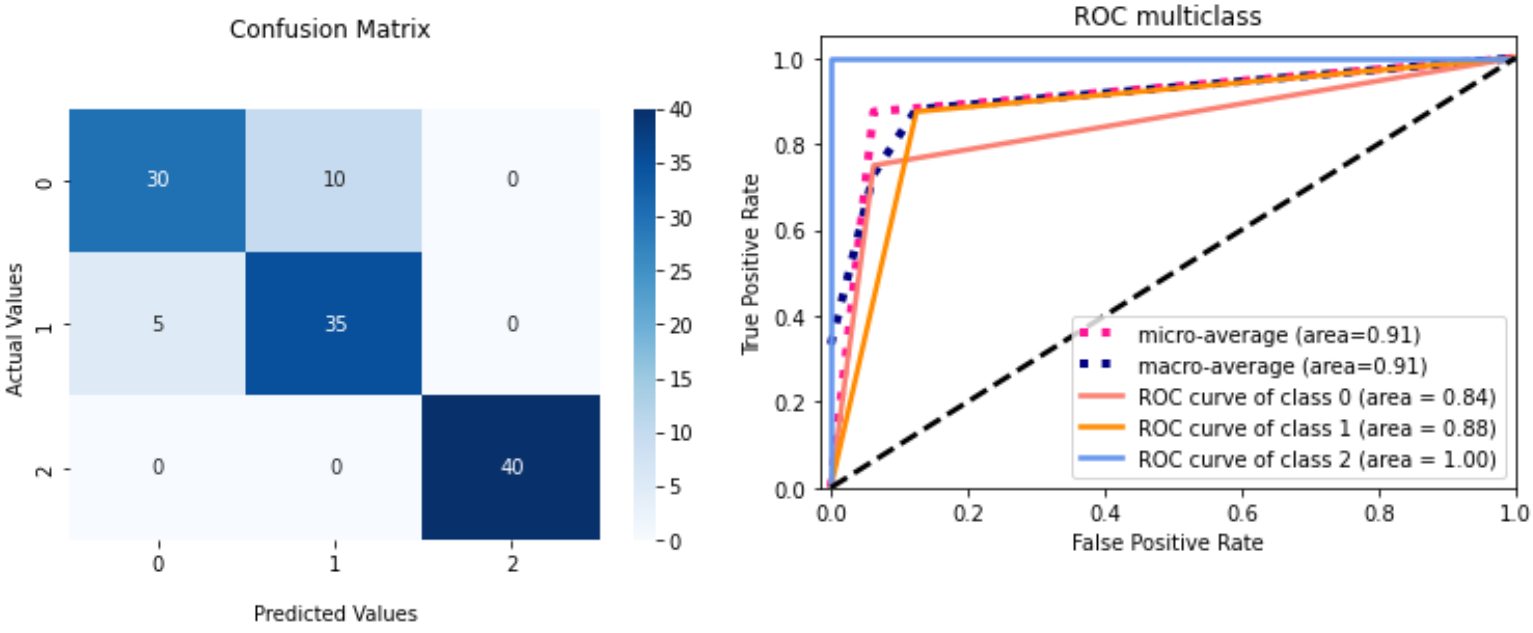


Figure 4-11 ROCAUC Curve/ Confusion Matrix of optimal training fold for the AD vs.CN vs. sMCI task.

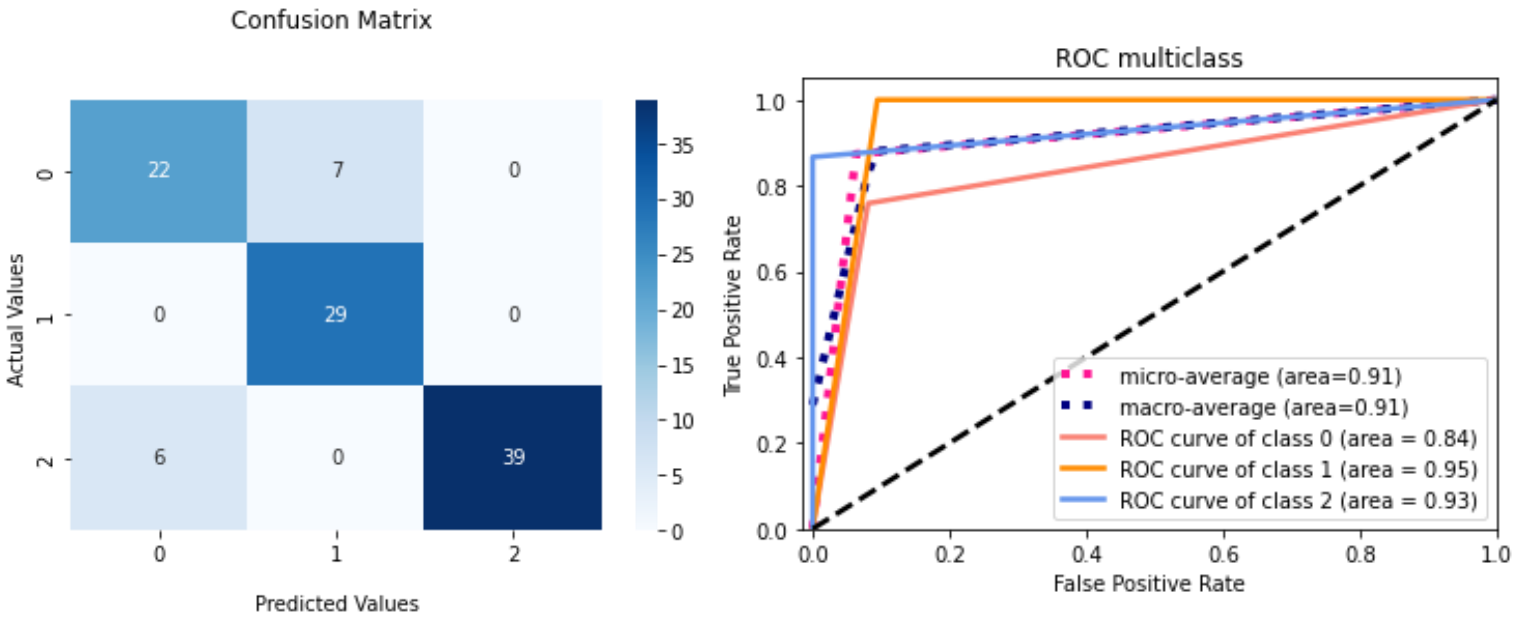


Figure 4-12 ROCAUC Curve/Confusion Matrix of optimal testing fold for the AD vs.CN vs. sMCI task.

4.4 Comparison with the Current state-of-the art methods

We compared our categorization findings to those reported in the literature, as shown in **Tables 4-2 and 4-3**. Most published literature regarding binary classification tasks utilized accuracy, sensitivity (SEN), specificity (SPE), balanced accuracy (BA), and AUC to demonstrate. We evaluated accuracy and AUC during our experiments. By using a confusion matrix of the best fold of the testing, we also computed SEN, SPE, and BA through the following formulas: TP/ (TP + FN), TN/ (TN + FP), and (SEN + SPE)/2, respectively.

The issue of whether patients with MCI can accurately self-diagnose their risk of developing AD remains essential to the development of viable treatments for the disease. Categorizing AD and sMCI is more challenging due to the subtler morphological changes that must be noticed, as demonstrated by the fact that the accuracy of several of the study results barely reached 70–80%.

The best Level 1 learning model, reported by Suk et al. [85], had a maximum accuracy of 74.82%. It is crucial to highlight that Level 1 approaches alter spatial localization in the feature extraction process of brain imaging data, as they rely on manual feature extraction. Without taking spatial relationships into account, it is hard to guess how the model decides to classify something in a reliable way. The Level 2 model proposed by Pan *et al.* [206] showed a maximum accuracy

of 83.81%. They suggested a multi-view separable pyramid network (MiSePyNet), a 2D CNN model that utilizes 18F-FDG PET images. MiSePyNet was built on the concept of quantized convolution and used independent slice and spatial-wise CNNs for each view. However, this Level 2 research only used a small part of the original datasets, thus disposing of any obvious outliers and making it hard to fairly compare its performance. In another study [140] carried out by Basaia *et al.*, 86.30% accuracy was obtained using a 3D CNN. MRI scans were segmented to create GM, WM, and CSF tissue probability maps in the MNI space. It was also built on a ROI-focused strategy rather than E2EL. Other studies [146] [144] [209] that used 2D TL with a pretrained network or local TL by transferring the knowledge of the AD vs. CN task to predict early diagnosis of AD obtained accuracies up to 82%.

DL Model	ACC	AUC	SPE	SEN	BA	Ref.
3D CNN [EfficientNet-B0]	**93.10**	**93.00**	**86.20**	**1.00**	**93.10**	*
3D CNN [DenseNet264]	82.50	82.50	82.50	82.50	82.50	[101]
Sparse Regression + 2D CNN	74.82	75.39	78.82	70.93	74.87	[85]
3D CNN	72.5	74.60	82.5	61.0	71.75	[15]
CAE + 3DCNN	73.95	79.11	70.71	77.46	74.08	[141]
3D CNN/ROI Based	86.3	-	88.7	84.0	86.35	[140]
3D CNN	76.0	81.0	76.0	71.0	73.5	[139]
2D CNN	83.81	88.89	87.50	75.76	81.63	[206]
MDNN/2D CNN	81.55	-	73.33	83.83	78.58	[144]
AlexNet + SVM 2D CNN	78.56	-	77.63	91.02	84.32	[146]
MM-SDPN	78.88	-	86.81	68.04	77.42	[209]

Table 4-2 : Matching up the findings of the AD vs. SMCI task with the results of the state-of-the-art DL models

* Proposed

DL Model	ACC	AUC	Precision	Recall	F-1 Score	Ref.
3D CNN [EfficientNet-B0]	**87.38**	**91.0**	**86.38**	**87.51**	**86.43**	*
CaffeNet/2D CNN	87.00					[19]
GogleNet/2D CNN	83.20					[19]
3D CNN	64.81	55.5	44.66		41.88	[121]

Table 4-3 : Matching up the findings of the AD vs. CN vs. SMCI task with the results of the state-of-the-art DL models

Only two research articles regarding multiclass categorization tasks could be found. One by Wu *et al.* [19] utilized 2D MRI slices and the pre-trained 2D CNN networks CaffeNet and GoogleNet, obtained an average accuracy of 87.00% and 83.20%, respectively. However, their implementation was based on Level 2 learning, and only obtained a 72.04% (for CaffeNet) and a 67.03% (for GoogleNet) accuracy rate for the classification of sMCI cases. Using Level 3 E2EL,MRI images as input and a basic 3D CNN model, Tufali *et al.* [121] conducted experiments for multiclass classification, but only obtained an average accuracy of 64.33% and an MCI class accuracy of 51.25%.

We achieved an accuracy of 93.10% in the evaluation of unseen data for the binary classification task and 87.38% for the multiclass classification task. This is significantly better than the early AD prediction accuracy reported by state-of-the-art methods in the last five years. Although our models are suitable to use in clinical settings to aid neuroradiologists, further training with more high-quality MRI scans from a diverse range of sources is required to ensure reproducibility. The scripts for model implementation for both types of tasks, with results, confusion matrices, and AUC graphs for every fold, can be accessed by using the link given in **Appendix B**.

4.5 Conclusion

Author did the fusion of E2EL with TL to improve the accuracy of the AD vs. sMCI task and used E2EL for the AD vs. CN vs. sMCI task. It requires the fine tuning of hyperparameters, and an appropriate **3D CNN architecture specifically designed for TL with excellent potential for generalization**; additionally, **MRI scans must be thoroughly pre-processed to maintain the spatial link and enhance image quality.** The results obtained in our experiments utilizing the **ADNI and IXI** datasets demonstrated that our models are more effective and efficient than the current state-of-the-art models for both binary and multiclass tasks.

Chapter 3 and Chapter 4 fulfill CONT#2 of this dissertation, which is the **identification and implementation of the optimum DL model. The** models (**DenseNet264 for the AD vs. CN task with E2EL; EfficientNet-B0 for the AD vs. sMCI task with E2EL and TL Fusion; EfficienNet-B0 for the AD vs. CN vs. sMCI task with E2EL)** were identified and will be used to **develop the ensemble learning-based DEEP-AD tool for achieving Goal #3 in Chapter 5**.

Chapter 5: DEEP-AD [Web tool for early detection of Alzheimer's]

Summary: As shown in **Figure 1-1 and Figure 1-2**, this chapter is about the third contribution, "**DEEP-AD," a web-based tool to aid neuroradiologists in their clinical judgment regarding the early detection of AD**," intended to help achieve the third goal" **A tool that neuroradiologists may use at their convenience to evaluate the model's robustness** ."

This chapters entails the creation of a **system prototype, DEEP-AD (CONT#3)**. The system prototype helped to (1) **fulfill the requirements** identified as essential to supporting neuroradiologists; (2) **show the viability of the ensemble of models (i.e., EficientNet-B0, DenseNet264)** and (3) **assess their usefulness in real-world scenarios (reported in chapter 6)** . In addition, the design of the interfaces was modified throughout the development process in accordance with the technologies employed.

5.1 Introduction

Deep-AD is a deep learning-based application for the early detection of AD. Built using open-source technologies streamlit[81], VScode[82], deployed on HuggingFace[83] Spaces.

The application has a reliable backend that is provided by state-of-the-art DL models EfficientNet-B0 and DenseNet264, which have already been identified, trained, and validated as described in chapters 3 and 4. The application can help doctors analyze AD with a low cost of computing.

We propose an ensemble learning approach to accurately determine the diagnostic status of AD, sMCI, and CN individuals based on T1W MRI scans. We utilize multiple DL models with a robust verification pipeline to enhance the results and increase the accuracy of the diagnosis. The ensemble learning method we propose shows the potential of fusing multiple models to improve the overall performance and can be applied to various other medical image classification tasks. Deep-AD is designed to be user-friendly and accessible, making it a useful tool for neuroradiologists and researchers alike.

[81] https://streamlit.io/
[82] https://code.visualstudio.com/
[83] https://huggingface.co/

In **Figure 5-1**, a use-case diagram of the tool has been given. The neuroradiologist must upload the MRI scans either in NIfTI format or DICOM format. After that, MRI will be preprocessed (bias field correction, denoising, brain extraction, and registration), and the preprocessed MRI will be entered as an input in the predict function, which will include an ensemble of DL models to classify it into different classes.

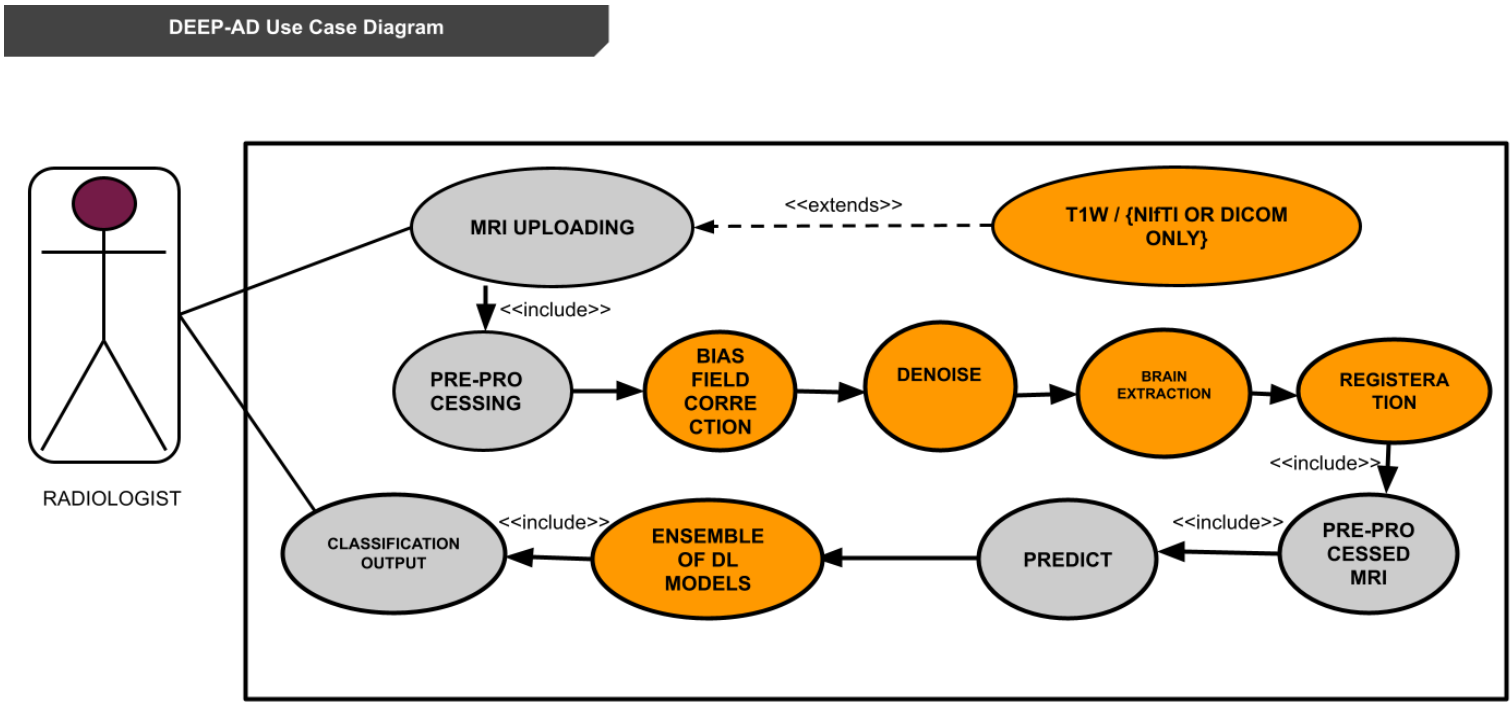


Figure 5-1 Use Case Diagram of DEEP-AD

5.2 Methodology

The **Ensemble learning**[84] method is applied through a verification pipeline. Ensemble learning combines many distinct models to improve generalization. Currently, DL architectures provide superior performance than shallow or conventional models. Deep ensemble learning models combine the benefits of both deep learning and ensemble learning, resulting in a model with superior generalization performance [227].

[84] https://machinelearningmastery.com/ensemble-methods-for-deep-learning-neural-networks/

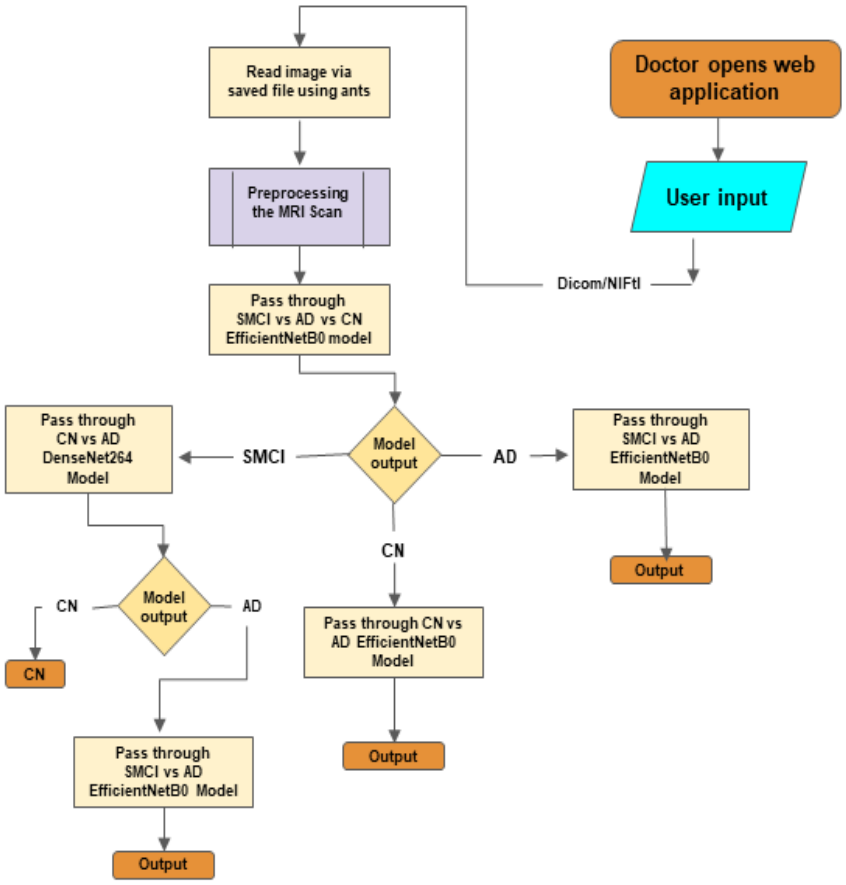
In **ensemble, the following three models** are utilized:

- **Base Model:** EfficientNet-B0, the output of the implementation presented in Chapter 4 (trained with E2EL, validated, and tested) for the multi classification task of AD vs. CN vs. sMCI.
- **Model 1:** EfficientNet-B0, the output of the implementation presented in Chapter 4 (trained with E2EL and TL, validated, and tested) for the binary classification task of AD vs. sMCI
- **Model 2:** DenseNet-264, the output of the implementation and comparative analysis presented in Chapter 3 (trained with E2EL, validated, and tested) for the binary classification task of AD vs. CN.

The base model is used to first process the incoming MRI scan. The input scans may be classified accurately and effectively by using this model in the first prediction step. If the base model predicts that the class is AD, then Model 1 will be used to look at the input scan to see if the diagnosis is sMCI or AD. If the base model predicts that the class is CN, model 2 will be used to find out if the diagnosis is CN or AD.

If the base model predicts the class as sMCI in the third scenario, the input scan will be checked via two separate pipelines since sMCI may be near CN (early MCI) or may be close to AD (late MCI or converted MCI). Model 2 is first used to analyze the input scan. If the outcome is CN, the output will be marked as in between CN and sMCI (early MCI or CN depending on the clinical findings); otherwise, the input scan will be analyzed by Model 2, if outcome is AD, the output will be marked as in between sMCI and AD (late MCI or converted MCI - depending on the clinical findings). The accuracy of the intermediate AD status is ensured by this multi-level verification technique and helps to minimize the chance of misdiagnosis. If both models produce the same results, that will be considered the final result.

Flowchart depicted in Figure 5-2 shows the evolution of user input from the first step to the last stage, to clearly describe the application's process flow. The ensemble learning algorithm, (**Algorithm 5**) which is the implementation of this flowchart is also shown.



Algorithm 5 DEEP-AD : Classifier

Require: T1W MRI scan either in NIfTI or DICOM format
Ensure: Classification in any of one AD/sMCI/CN .

```
Scan ← Input MRI Scan                          ▷ Either in NIfTI or DICOM format
Scan ← Bias Field Correction(Scan)             ▷ Prepossessing step
Scan ← Denoising(Scan)                         ▷ Prepossessing step
Scan ← Brain Extraction(Scan)                  ▷ Prepossessing step
Scan ← Registration(Scan)                      ▷ Prepossessing step
Output ← Base Model(Scan)                      ▷ Prepossessing step
if Output is AD then
    Output ← Model1(Scan)
else if Output is CN then
    Output ← Model2(Scan)
else if Output is sMCI then
    Output ← Model2(Scan)
    if Output is AD then
        Output ← Model1(Scan)
    end if
end if
```

Figure 5-2 Flow chart & Algorithm of DEEP-AD

5.3 Frameworks, Tools, and IDEs

In the development of the Deep-AD application, we utilized open-source tools VS Code and Streamlit for coding and building the user interface. These tools provided us with the flexibility and ease of use, necessary to create a user-friendly application.

Additionally, we deployed the application on HuggingFace Spaces, a platform that enables easy deployment of machine learning models. This deployment process allowed us to make the application accessible to a wide range of users, including neuroradiologists and researchers. Furthermore, the use of HuggingFace Spaces reduced computing costs and increased efficiency. Overall, the combination of these open-source tools and platforms helped us create a reliable and efficient application. The pros of using these open sources are listed below.

- **Streamlit** was a crucial tool that assisted us in creating the user interface. This open-source library allowed us to build a visually pleasing and intuitive interface that makes it easy for the users to input and process MRI scans.

  Streamlit's user-friendly interface made the process of creating the application's layout and navigation simple and efficient. The interface was designed to enable the users provide the inputs of the scans, process them with the ensemble learning algorithm, and get the results in a clear and easy-to-understand format.

  The use of Streamlit greatly assisted in creating an application that is not only accurate and efficient but also user-friendly, which is crucial for a medical application like Deep AD.

- In the design and implementation of the Deep AD application, we utilized HuggingFace Spaces as the deployment platform. This platform provided us with the capability to deploy the application with ease, making it accessible to a wide range of individuals, including neuroradiologists and researchers.

  Additionally, the platform enabled us to decrease the computational costs required to run the application, which is a crucial consideration in a medical application.

  The platform also provided us with an intuitive interface for the monitoring and management of the models, which greatly aided in the deployment process and ensured the smooth functioning of the application.

- VS Code played a significant role as the primary tool for coding and debugging. Its user-friendly interface, built-in debugging features, and various useful extensions made the coding process efficient and streamlined.

  VSCode's integration with Git allowed us to effectively manage and track the changes made to the code throughout the development process. Additionally, the built-in debugging features of VSCode helped us identify and fix any bugs that arose during the development process, ensuring the smooth functioning of the application. The use of VSCode also allowed us to write and test the code for the application in an organized and efficient manner.

5.4 Prototype

The design and implementation of the prototype for Deep AD were executed utilizing open-source technologies. The goal of the prototype was to create a user-friendly and easily accessible tool. The prototype was developed in several phases, beginning with the selection of the appropriate DL models and their deployment on the HuggingFace Spaces platform. Prototype can be accessed by using the website https://huggingface.co/spaces/Deevyankar/Deep-AD.

The screen shot of the home page is shown in **Figure 5-3**. The platform on which the application was deployed was designed to provide easy access and keep computational costs low. We must pay 0.03 $ /hour to use an 8-v CPU and 32 GB of RAM, which are sufficient to run this app. The prototype exemplifies the potential of deep AD to serve as a crucial tool for the early detection of AD.

5.4.1 Guidelines for Interpreting Results

To aid users in comprehending the result, we have created the following guidelines:

- The first rule is that a choice may be regarded as final if both models provide the same result (i.e., "no conflict").
- Second Rule: If the results from the two models are different, the **Table 5.1** below provides a variety of options and factors to think about.

Neuroradiologists may utilize these rules in conjunction with clinical evidence to reach a conclusion.

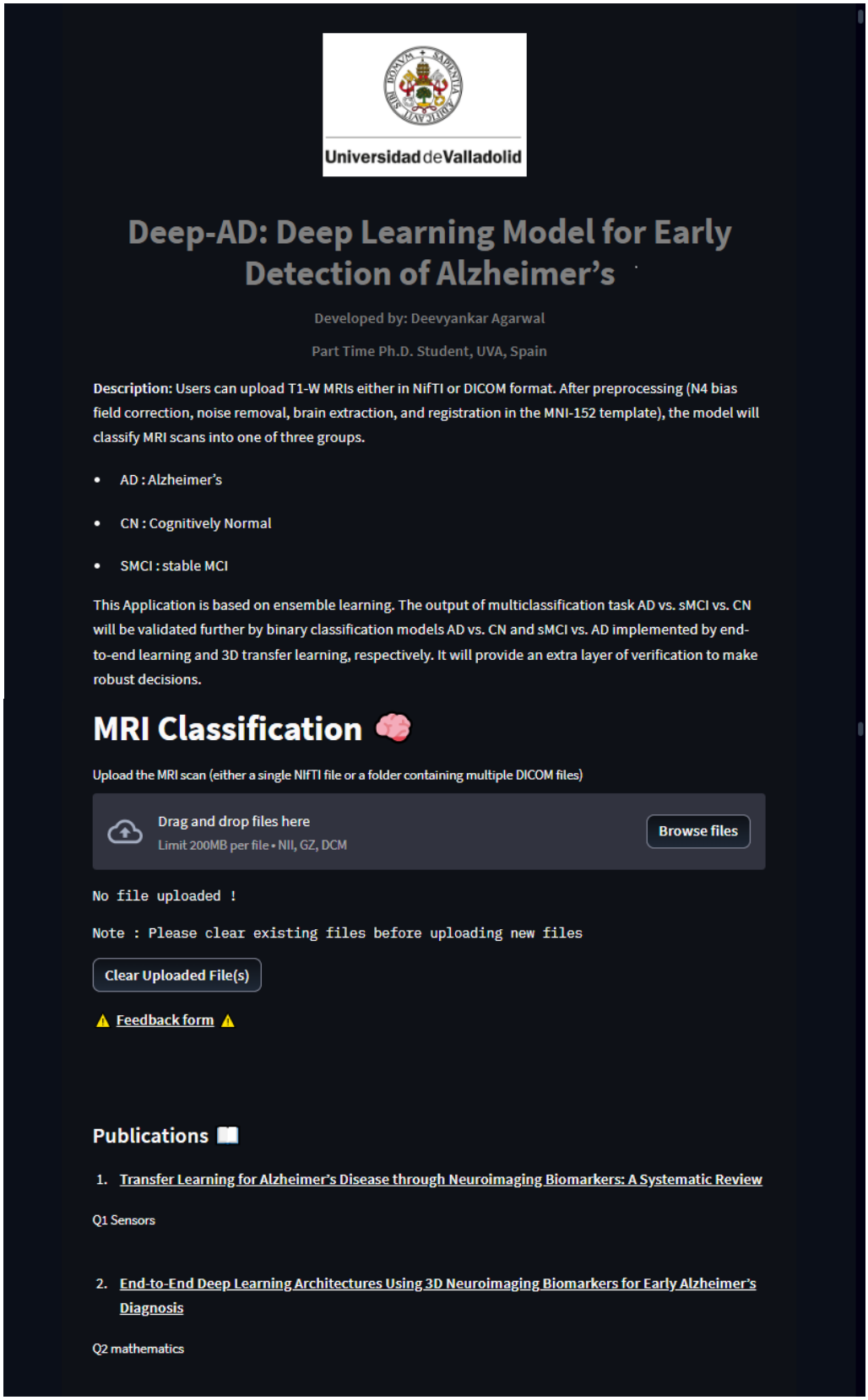


Figure 5-3 Home Page Screen Shot

Base Model Prediction	Second Prediction	Suggestion-1 (S-1)	Suggestion -2 (S-2)
AD/sMCI	sMCI/AD	Converter MCI or Late MCI	The second model's forecast is definitive since it performed better in both testing and training.
CN	AD	AD	AD
CN/sMCI	sMCI/CN	Early MCI or stable MCI	The second model's forecast is definitive since it performed better in both testing and training.

Table 5-1 Guidelines for interpreting the results of DEEP-AD

5.5 Conclusion

We created **DEEP-AD in this chapter as a contribution toward attaining Goal #3**. By using the **actual data set of Spanish individuals, the qualitative analysis of DEEP-AD is reported in Chapter 6.**

DEEP-AD has the potential to revolutionize the field by providing an objective and efficient means of diagnosis. It is designed to be user-friendly and accessible, making it a useful tool for clinicians and researchers alike. The proposed approach has a good chance of making it easier to find and diagnose AD early and can help come up with better plans for how to treat it.

It utilizes an ensemble learning approach that incorporates multiple state-of-the-art DL models and a robust verification pipeline to increase the accuracy of the diagnosis. **The application is user-friendly, easy to access, and efficient** because it uses open-source technologies and is built on a platform that makes it easy to get to and reduces the cost of computing.

Chapter 6: DEEP-AD validation utilizing MRI scans of Spanish subjects by neuroradiologist

Summary: As shown in Figure 1-1 and Figure 1-2, this chapter is about the third contribution, "**DEEP-AD," a web-based tool to aid neuroradiologists in their clinical judgment regarding the early detection of AD**," intended to help achieve the third goal" **A tool that neuroradiologists may use at their convenience to evaluate the model's robustness** ."

In this chapter, we (1) **assess the DEEP-AD's performance and usability**, (2) **confirm the system's level of support for the established criteria**, and (3) monitor the effect on people. Given this context, a data set of 41 MRI scans was used, which was provided by neuroradiologists of HT Medica . **Dr. Alvero Berbis, Head of the Neurology Department at HT Medica in Spain, validated the DEEP-AD to verify the predictions of Spanish subjects (41 MRI scans) and obtained an accuracy of 82.90%**. His feedback gave us useful information that lets us judge the success of the linked dissertation goals in a positive way.

6.1 Introduction

The advent of DL-based approaches has brought about a paradigm shift in the realm of early detection of AD. The utilization of such sophisticated techniques has the potential to revolutionize the field by providing a quantitative and efficient means of diagnosis. However, it is imperative to critically evaluate the performance of these applications to ascertain their clinical utility. In this dissertation, we aimed to systematically evaluate the performance of DEEP-AD, by utilizing neuroradiologists as evaluators.

The evaluation design employed in this dissertation for assessing the performance of DEEP-AD is of paramount significance in determining the clinical utility of the application. The design adopted in this study is a retrospective design [228].A dataset comprising of Spanish subjects was utilized to evaluate the performance of Deep AD.

The results of the evaluation revealed that DEEP-AD had 82.90% accuracy. The diagnostic outcomes produced by the application were in concordance with the clinical diagnoses obtained from the neuroradiologists in the 34 cases out of 41.

The report provided by Dr. Alvero Berbis for this verification procedure is included in **Appendix C**, along with his suggestions. This serves as evidence that the application has the potential to provide reliable diagnostic outcomes.

6.2 Methodology

The methodology applied is shown in Figure **6-1**. It consist of the following steps.

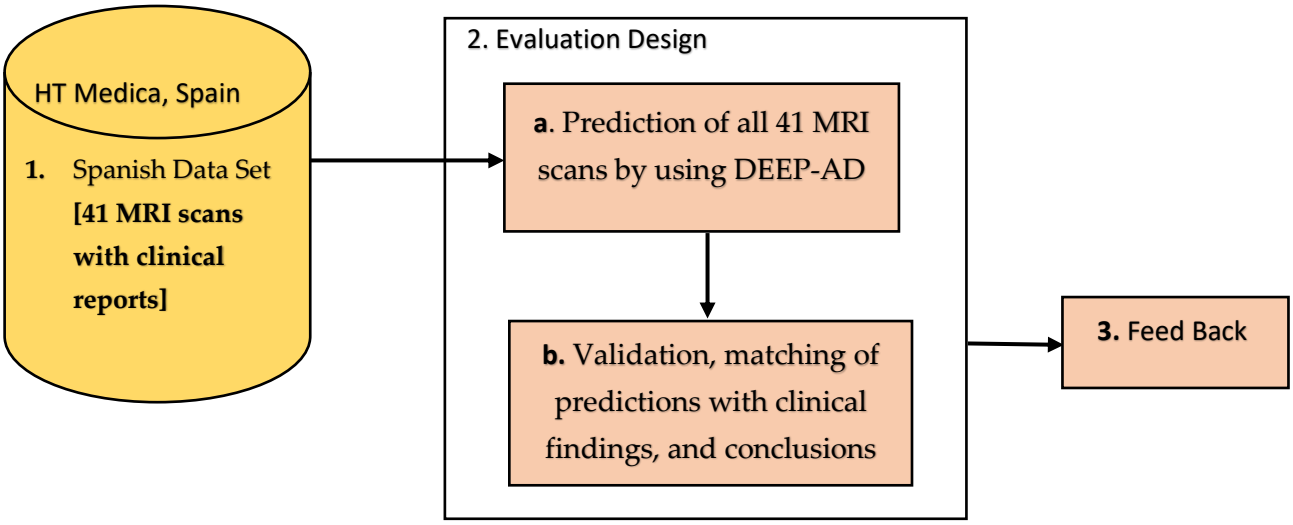


Figure 6-1 DEEP AD's Evaluation Design, Centered on a Neuroradiologist Perspectives, and a Spanish Data Set

6.2.1 Dataset of Spanish subjects

Obtained from HT Medica, Spain. The dataset comprises of a heterogeneous population of individuals with varying degrees of cognitive impairment. The sample population is procured from multiple centers, providing a representative and diverse sample reflective of the Spanish population. It contains clinical reports and T1W MRI scans of 41 individuals with AD, CN, and MCI. The following details are included in reports.

1. Scan date
2. Location of the hospital in Spain
3. Clinical information
4. Findings
5. Conclusion
6. Name of the neuroradiologist with signature and collegiate number.

Figure 6-2 shows sample of one report. The research's findings and conclusions were utilized to confirm DEEP-AD predictions.

Users can access all the reports and MRI scans. **Other researchers at UVA may use these datasets in their master's or doctoral dissertations. [can be shared upon reasonable request]**

Fecha informe: 17/02/2022

Realizado en: HT SERCOSA
Solicitado por

Paciente:

EXPLORACIONES:
RMI de cráneo

INFORMACION CLINICA
Deterioro cognitivo de perfil fronto-temporal/alzheimer.

COMPARADO CON:
TAC del 16/11/2015

HALLAZGOS:
RM de craneo

Acentuación de surcos corticales, cisternas silvianas y sistema ventricular en relación a involución cerebral mixta.
Minima hiperintensidad de señal periventricular compatible con incipiente leucoaraiosis.
Cuerpo calloso y charnela craneooccipital sin evidencia de alteraciones morfológicas.
Órbitas, globos oculares, tejido graso retro ocular y nervios ópticos sin particularidades.
Región selar y pineal sin particularidades.
En secuencias de difusión no se observaron imágenes con efectos restrictivos.

CONCLUSION:
RM de craneo

Signos de involución cerebral mixta.
Incipiente leucoaraiosis periventricular.

Fdo: Dr. Nicolas Pagni
Nº Colegiado: 7221

CONSENTIMIENTO

El paciente ha realizado el consentimiento para la realización de la exploración de forma verbal.

Figure 6-2 Sample of clinical report provided by HT Medica, Spain

6.2.2 Evaluation Design

The evaluation design employed for assessing the performance of DEEP-AD is of paramount significance in determining the clinical utility of the application. The design adopted in this study is a retrospective design, it contains following 2 steps.

a. The application is used to classify the MRI scans of Spanish subjects.

b. The results were compared to the clinical diagnoses by neuroradiologists.

The neuroradiologists were provided with the diagnostic outcomes produced by DEEP-AD for Spanish datasets, guidelines for interpreting the results and were asked to independently verify the results. This allows for the assessment of inter-rater reliability of the application and provides insight into the agreement between the diagnostic outcomes produced by the application and the clinical diagnoses obtained from the neuroradiologists.

Furthermore, the use of neuroradiologists as evaluators in this study is of particular significance, as they are experts in the field of neuroimaging and have extensive experience in the diagnosis of AD. This allows for a more robust evaluation of the application's performance.

The tool was created with the intention of being user-friendly, so that even neuroradiologists who are not technologically savvy can easily use and comprehend it. The program generates diagnostic results in a graphical form that is simple to interpret and neuroradiologists can easily correlate to the clinical findings. This makes it an ideal tool for both neuroradiologists and researchers.

6.2.3 Feedback

The following are the most significant points.

- Most of the reports were found **(34 out of 41)** to agree with the clinical diagnoses.
- The interface was found to be user-friendly and easy to use.
- The application provides clear diagnostic outcomes, and the results are easy to interpret.
- The application is cost-effective and easy to adopt.
- DEEP-AD needs further training using data from several hospitals to improve its accuracy and generalizability.
- Seek out the factors that seem to have had the most significant role in the decision-making process.

6.2.3 Results

This part focused on the input scans and the application's output predictions, which are shown in **Table 6-1**. Using the defined standards for evaluating the outcome of AD, the output predictions produced by Deep-AD for each input scan were compared to the clinical diagnosis by a neuroradiologist.

Following are the three outputs for each MRI scan's execution through DEEP-AD.

1. Display of input MRI Scan
2. MRI scan after preprocessing
3. Graphical Output of both the models.

The execution result of various MRI scans from Spanish datasets that were classified using DEEP-AD are shown below. We tried to address all potential scenarios.

Case 1: [No Match] Contradiction with clinical results : Both models give same output.

One of the input scans [id 1] clinical reports reported a patient with frontotemporal dementia or Alzheimer's, mixed cerebral involution, and periventricular leukoaraiosis incipient. The patient was recognized as CN through both models of Deep-AD. **Figures 6-3, 6-4, and 6-5** display all three of the execution's outcomes.

Case 2 : [Match] Exact match with clinical results : Both Models give same output.

Another input scan [id 43] clinical reports reported a patient with Incipient frontal cortical subcortical atrophy. Periventricular leukoaraiosis and focal ischemic gliosis-type lesions in the supratentorial and infratentorial white matter of the brain (Rule out Alzheimer's). The patient was recognized as CN through both models of Deep-AD. **Figures 6-6, 6-7, and 6-8** display all three of the execution's outcomes.

Case 3: [Match] Both models give different Output : Decision according to S-1 of Table 5-1

A third input scan **[id 24]** clinical reports reported a patient with changes due to Acute ischemic focus in the left thalamic region .Leukoaraiosis. Signs of involution. Cognitive impairment type Alzheimer.

Since one model's output prediction was sMCI and another was CN, the final forecast based on S-1 was early MCI; nevertheless, a neuroradiologist might pick it as sMCI based on the clinical

history and symptoms. *Thus, our tool and the clinical findings may help together with the decision-making process.* **Figures 6-9, 6-10, and 6-11** display all three of the execution's outcomes.

Case 4: [Match] Both models give different Output : Decision according to S-1 of Table 5-1

A fourth input scan **[id 47]** clinical reports reported a patient with possible mild Alzheimer's disease and depressive symptoms. Since one model's output prediction was AD and another was sMCI, the final forecast based on S-1 was converter MCI/late MCI, neuroradiologist further can choose anyone based on the clinical symptoms. **Figures 6-12, 6-13, and 6-14** display all three of the execution's outcomes.

The results of this evaluation provide valuable insights into the performance of Deep-AD and its potential to improve diagnostic processes and benefit patients.

Deep-AD Performance

SNo.	Id	Report date	Conclusion	Clinical Information	Deep-AD Prediction	Match?	Remark
1	1	2/17/2022	Signs of mixed cerebral involution, Incipient periventricular leukoaraiosis	Cognitive impairment of the fronto-temporal profile/Alzheimer's	**CN**	**No**	
2	3	1/4/2022	The cortical and central involutional changes with volumetric decrease of both hippocampi in this period of observation have progressed with respect to the previous study.	Cognitive impairment in a moderate stage of probable neurodegenerative origin.	**AD/sMCI**	**Yes**	**S-1**
3	4	13/11/2021	Moderate corticosubcortical atrophy. We observed mild ventriculomegaly, especially at the expense of the lateral ventricles, probably due to associated subcortical atrophy.	Mild-moderate cognitive impairment with amnesic predominance of probable neurodegenerative origin.	**AD**	**Yes**	
4	7	7/9/2021	Signs of mixed cerebral involution associated with foci of gliosis and leukoaraiosis.	68 years. Probable mild-moderate Alzheimer's disease	**CN**	**No**	
5	8	8/24/2020	Lacunar foci of gliosis-ischemia in bilateral and subcortical periatrial white matter of the parietal lobes. Moderate generalized cortico-subcortical atrophy, without locoregional predominance.	Alzheimer's principle.	**sMCI/CN**	**Yes**	**S-1**
6	13	9/12/2019	Chronic right occipital ischemic infarction. Multiple focal lesions on T2* gradient echo sequence. Mesial temporal atrophy.	Acute language deficit after which he presents subacute severe cognitive-behavioral impairment.	**CN**	**Yes**	

Deep-AD Performance							
SNo.	**Id**	**Report date**	**Conclusion**	**Clinical Information**	**Deep-AD Prediction**	**Match?**	**Remark**
7	14	23/10/2019	We identified a moderate generalized cortical subcortical atrophy, predominantly posterior parietal with increased furrows at the level of the convexity.	. Possible Alzheimer.	**AD/sMCI**	**Yes**	**S-1**
8	15	6/22/2019	Signs of frontoparietal corticosubcortical atrophy. Atrophy of both hippocampi with frank left predominance. Periventricular leukoaraiosis and multiple ischemic-type lesions in white matter of both hemispheres	Cognitive impairment in a patient with suspected Alzheimer's.	**AD/sMCI**	**Yes**	**S-1**
9	18	6/9/2018	Bilateral frontotemporal corticosubcortical atrophy as described. Moderate-severe chronic microangiopathy lesions.	Cognitive decline.	**sMCI/CN**	**Yes**	**S-1**
10	24	5/15/2021	Acute ischemic focus in the left thalamic region .Leukoaraiosis. Foci of gliosis of probable ischemic origin. Signs of involution. Hydrocephalus.	Cognitive impairment type Alzheimer. Hydrocephalus	**sMCI/CN**	**Yes**	**S-1**
11	26	12/28/2019	Brain involution. Leukoaraiosis .Multiple foci of gliosis/sequel ischemic lesions of small vessels.	Parkinsonism and Alzheimer's	**AD/sMCI**	**Yes**	**S-1**
12	28	7/7/2018	Signs of moderate mixed cerebral involution with periventricular leukoaraiosis and multiple ischemic gliosis-type lesions in supra and infratentorial white matter of the brain.	Alzheimer-type cognitive impairment in a patient with a history of breast cancer.	**CN/AD**	**Yes**	**S-2**
13	47	3/27/2022	Few white matter lesions of chronic small vessel ischemic type. Abnormal atrophy for age with moderate-severe volume loss in both hippocampi, more advanced in the left. Assess primary degenerative process type AD.	He does not know what day we are in nor is he able to name the objects on the photo test, she doesn't want to go out.	**AD/sMCI**	**Yes**	**S-1**
14	48	2/14/2022	GENERALIZED CORTICO-SUBCORTICAL VOLUME LOSS WITH IMPORTANT MESIAL TEMPORARY INVOLVEMENT, ALZHEIMER'S DISEASE CANNOT BE RULED OUT.	mild cognitive impairment	**CN**	**No**	

Deep-AD Performance							
SNo.	**Id**	**Report date**	**Conclusion**	**Clinical Information**	**Deep-AD Prediction**	**Match?**	**Remark**
15	50	1/24/2022	Changes due to chronic cerebrovascular disease	CVRF: hypertension, moderate obesity, type II diabetes, dyslipidemia,	**AD/sMCI**	**No**	**S-1**
16	54	7/11/2021	Atrophy of the described characteristics, assess the possibility of a primary degenerative process of the Alzheimer type.	Patient with cognitive impairment suggestive of Alzheimer's disease.	**AD/sMCI**	**Yes**	**S-1**
17	55	11/2/2021	Mild changes due to chronic cerebrovascular disease. Atrophy of the described characteristics, assess primary degenerative process type AD.	Progressive cognitive impairment.	**CN**	**No**	
18	61	9/20/2021	Changes due to chronic cerebrovascular disease (Fazekas 1-2). Atrophy of the described characteristics, suspicious for an Alzheimer-type degenerative process.	A 79-year-old patient with dementia that could be due to Levy bodies or, less likely, Alzheimer's type.	**CN**	**Yes**	
19	62	9/14/2021	Moderate signs of diffuse cerebral atrophy with cortical predominance. Mild areas of gliosis and/or demyelination of chronic ischemic origin.	Alzheimer's.	**AD/sMCI**	**Yes**	**S-1**
20	64	3/9/2021	Possible ischemic lesions of small pontine vessels. Signs of cortical atrophy with slight parietal and temporal predilection	Male patient with cognitive impairment of amnesic type	**AD/sMCI**	**Yes**	**S-1**
21	65	2/4/2021	Signs of cerebellar atrophy and mild nonspecific brainstem signal alteration, to assess small vessel or degenerative ischemia.	They consulted for two episodes of short term generalized tonic-colonic seizures, one a month and a half ago and the other two and a half months ago,	**AD/sMCI**	**Yes**	**S-1**
22	67	2/12/2020	Small vessel ischemic disease evolved at the supratentorial level. Mild cortico-subcortical involution, somewhat more prominent in the bilateral temporal region.	A 66-year-old patient with cognitive impairment suspected of Alzheimer's but has multiple CVRF. A. familiar's dementia. Request cranial MRI	**AD/sMCI**	**Yes**	**S-1**
23	69	10/16/2020	Atrophy with special involvement of the two hippocampi, suggestive of a primary Alzheimer-type degenerative process. Discrete changes due to chronic occlusive vascular encephalopathy.	An 83-year-old patient under study for mental deterioration/dementia. Assess degree of atrophy.	**CN**	**Yes**	

Deep-AD Performance							
SNo.	Id	Report date	Conclusion	Clinical Information	Deep-AD Prediction	Match?	Remark
24	72	7/3/2020	Age-related involutional changes with hippocampal atrophy, findings that could be related to Alzheimer's disease, it is suggested to correlate with tests of cognitive and clinical function. Ischemic chronic supratentorial leukoencephalopathy	Cognitive impairment, movement disturbance.	**AD,/sMCI**	**Yes**	**S-2**
25	20	2/1/2019	Stability of findings with minimal growth of the intraventricular lesion in the left atrium.	Intraventricular meningioma	**AD/sMCI**	**Yes**	**S-1**
26	29	3/10/2022	Possible mild Alzheimer's disease. depressive symptoms	Possible mild Alzheimer's disease. depressive symptoms	**AD**	**Yes**	
27	30	2/5/2022	Mild signs of mixed cerebral involution.	possible senile Alzheimer's dementia	**sMCI/CN**	**Yes**	**S-1**
28	31	11/25/2021	Significant mesial and cerebral atrophy . Arteriolar vascular leukoencephalopathy (Fazekas type 1). No other alterations.	Mnesic failures under study. Multiple lacunar infarcts	**CN, AD**	**Yes**	**S-2**
29	32	9/28/2021	Small foci of gliosis-ischemia at the supratentorial level. Bilateral temporal atrophy.	Alzheimer disease.	**AD/sMCI**	**Yes**	**S-1**
30	33	10/20/2020	Atrophy with moderate bilateral hippocampal volume loss. The possibility of a primary Alzheimer-type degenerative process should be assessed. Doubtful lesion in the left middle fossa for which control with contrast is recommended if considered appropriate.	66-year-old patient under study for moderate cognitive impairment, possible Alzheimer-type degenerative origin.	**AD/sMCI**	**Yes**	**S-1**
31	34	5/31/2020	Cerebral atrophy with involvement of the Sylvian valleys and bilateral mesial temporal areas. Arteriolar vascular leukoencephalopathy (Fazekas type 1). No other alterations.	Alzheimer's disease	**CN**	**NO**	
32	35	02/27/2020	Moderate diffuse cerebral atrophy predominantly perisylvian and parietal (GCA 2) and mesial temporal, more evident left (MTA 2/3), compatible with end. Alzheimer's. Acute alteration is not visualized.	Alzheimer's type dementia. Vascular component? Acute worsening a month ago.	**AD/sMCI**	**Yes**	**S-1**

Deep-AD Performance

SNo.	Id	Report date	Conclusion	Clinical Information	Deep-AD Prediction	Match?	Remark
33	36	9/21/2019	Signs of moderate mixed cerebral involution with predominance of cortico-subcortical frontotemporal atrophy and right hippocampal atrophy.	Probable Alzheimer's disease.	**CN**	**NO**	
34	37	05/05/2019	Signs of moderate mixed cerebral involution.	Alzheimer-type cognitive impairment.	**AD/sMCI**	**Yes**	**S-1**
35	39	6/10/2018	Periventricular leukoaraiosis and multiple ischemic gliosis-type lesions in the white matter of both hemispheres.	Suspected Alzheimer's disease. Assess pattern of atrophy.	**AD/sMCI**	**Yes**	**S-1**
36	40	5/23/2018	Isolated Ischemic gliosis-type lesions in the white matter of both hemispheres.	Alzheimer's disease with c cerebrovascular accident and left hemiparesis. Chronic headache.	**AD/sMCI**	**Yes**	**S-1**
37	41	2/9/2022	Calcium deposits with bilateral and symmetrical distribution at the level of the basal ganglia and cerebellar dentate nuclei with non-specific characteristics.	A 68-year-old male with suspected progressive cognitive impairment with an amnesic profile. Suspected underlying neurodegenerative process such as Alzheimer's disease. Assess degree of cerebral atrophy.	**sMCI/CN**	**Yes**	**S-1**
38	42	26/08/2021	MRI of the skull Signs of mixed cerebral involution associated with foci of ischemic gliosis and leukoaraiosis. Left cerebellar ischemic sequela.	Cognitive impairment. Possible Alzheimer's.	**sMCI/CN**	**Yes**	**S-1**
39	43	1/4/2021	Incipient frontal cortical subcortical atrophy. Periventricular leukoaraiosis and focal ischemic gliosis-type lesions in the supratentorial and infratentorial white matter of the brain.	Rule out Alzheimer's.	**CN**	**Yes**	
40	44	30/07/2018	Acute-subacute ischemic lesions in the subcortical and insular right temporal and right parietal cortical regions (with probable subacute hemorrhagic transformation).	Left hemicorporeal deficit.	**sMCI/CN**	**Yes**	**S-1**

Deep-AD Performance							
SNo.	**Id**	**Report date**	**Conclusion**	**Clinical Information**	**Deep-AD Prediction**	**Match?**	**Remark**
41	46	9/5/2018	Severe, generalized cortico-subcortical atrophy of the temporal lobes, suggestive of Alzheimer-type degeneration.	Unrelated seizures with two nonspecific lesions in the left semioval center to rule out ischemic or metastatic disease.	**sMCI/CN**	**Yes**	**S-1**

Table 6-1 Clinical details and DEEP-AD prognosis for Spanish individuals' MRI scans

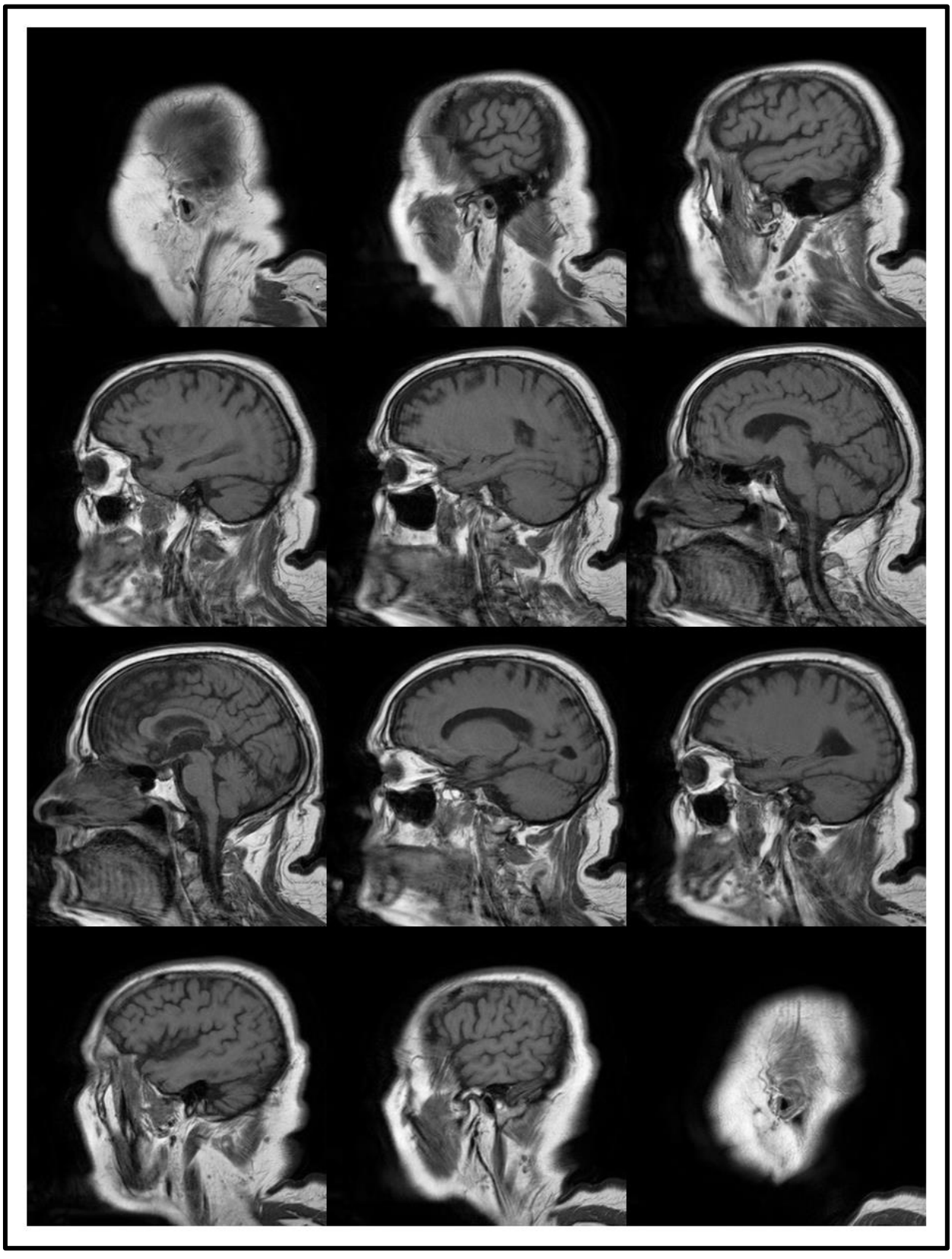


Figure 6-3 Case 1 MRI Scan Input

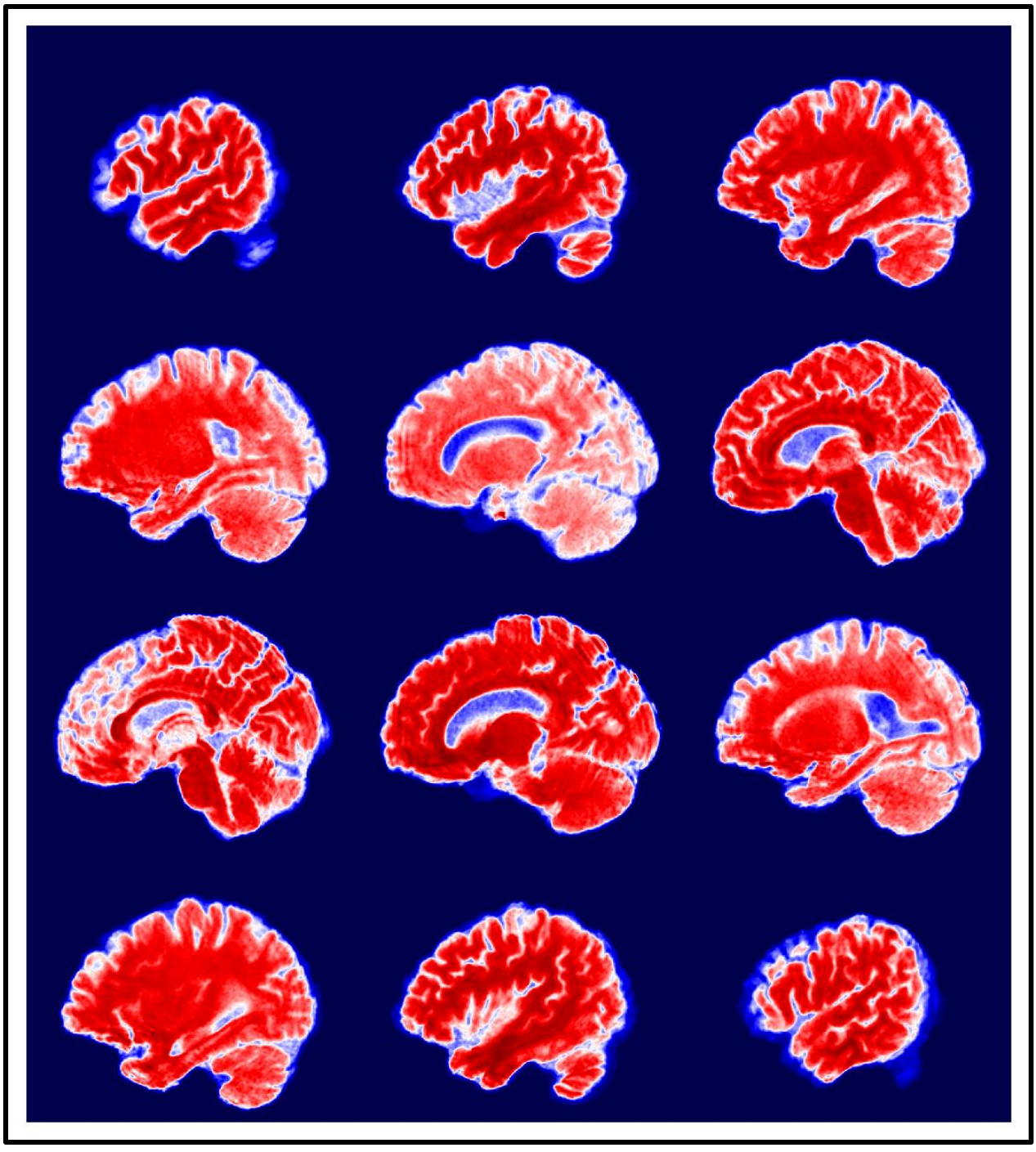


Figure 6-4 Case 1 MRI Scan After Preprocessing

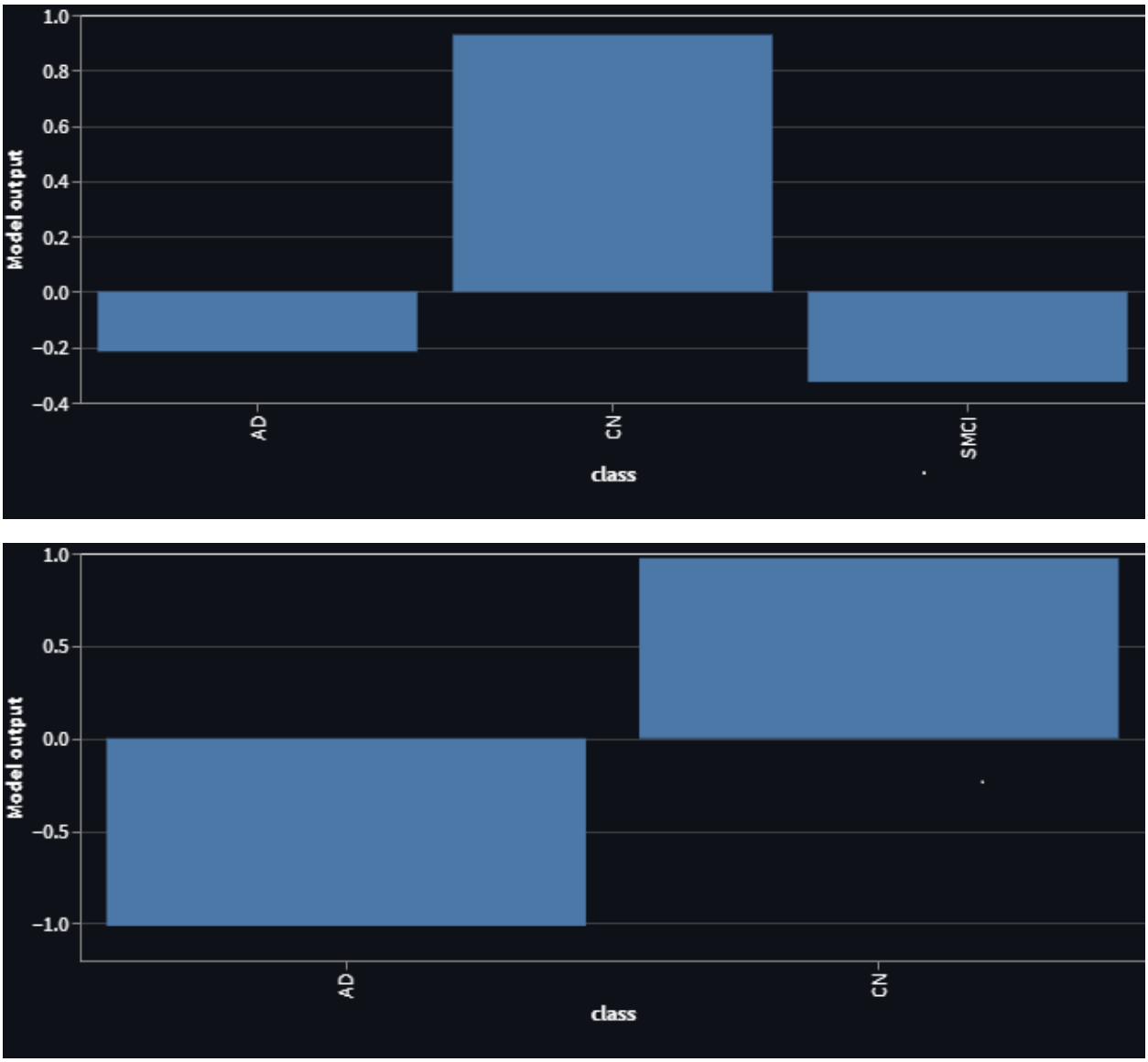


Figure 6-5 Case 1: Output of both the models

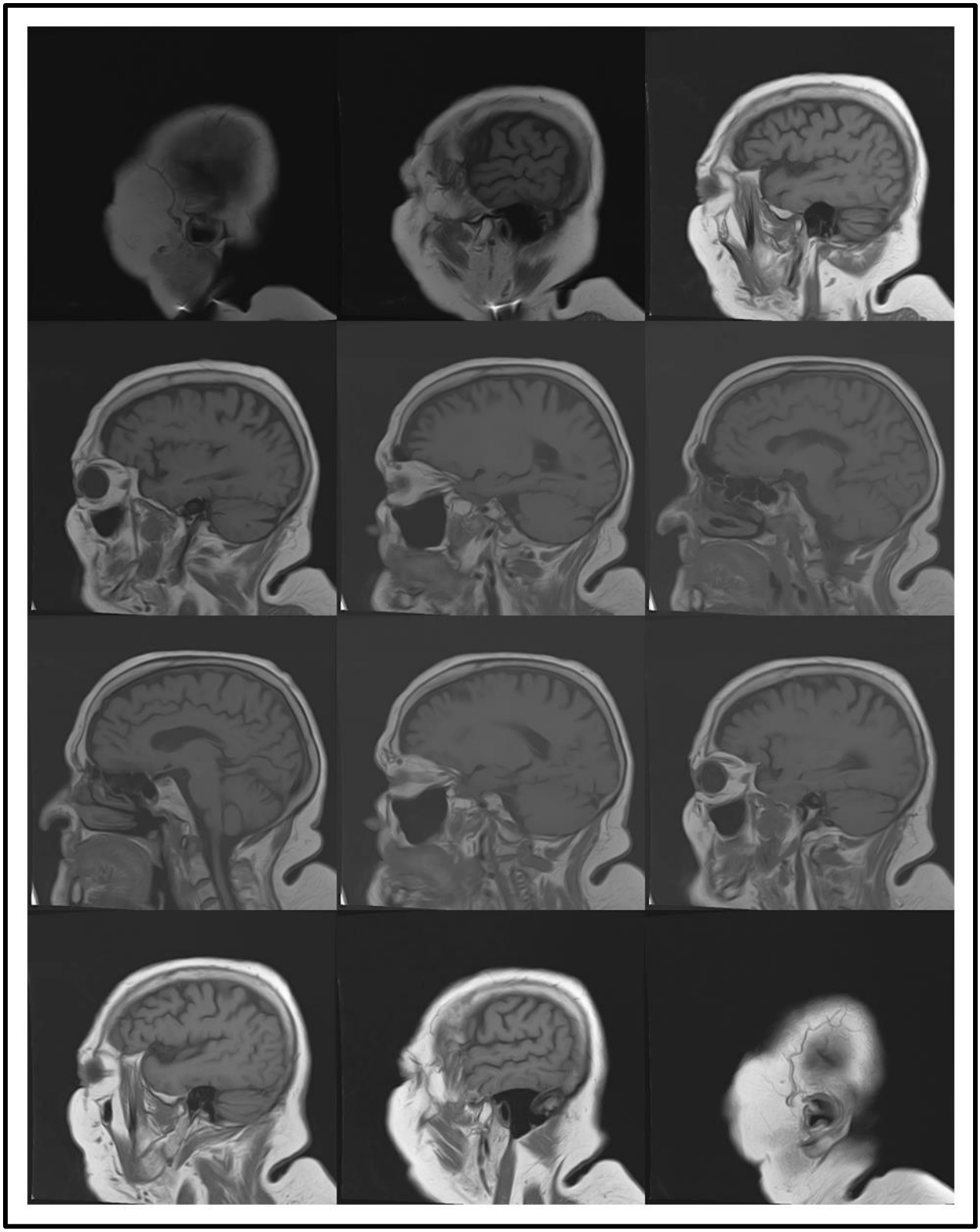


Figure 6-6 Case 2 MRI Scan Input

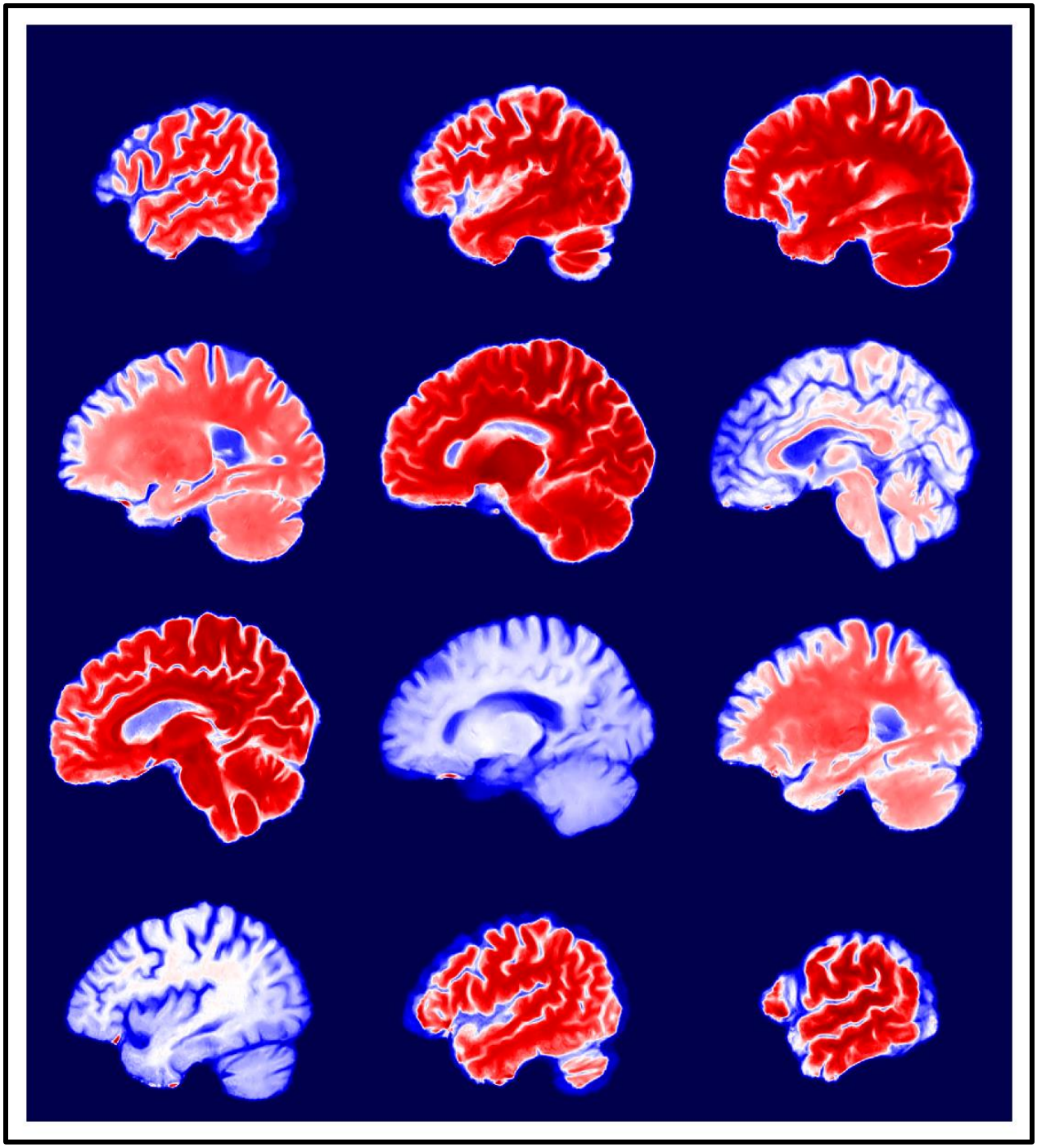


Figure 6-7 Case 2 MRI Scan After Preprocessing

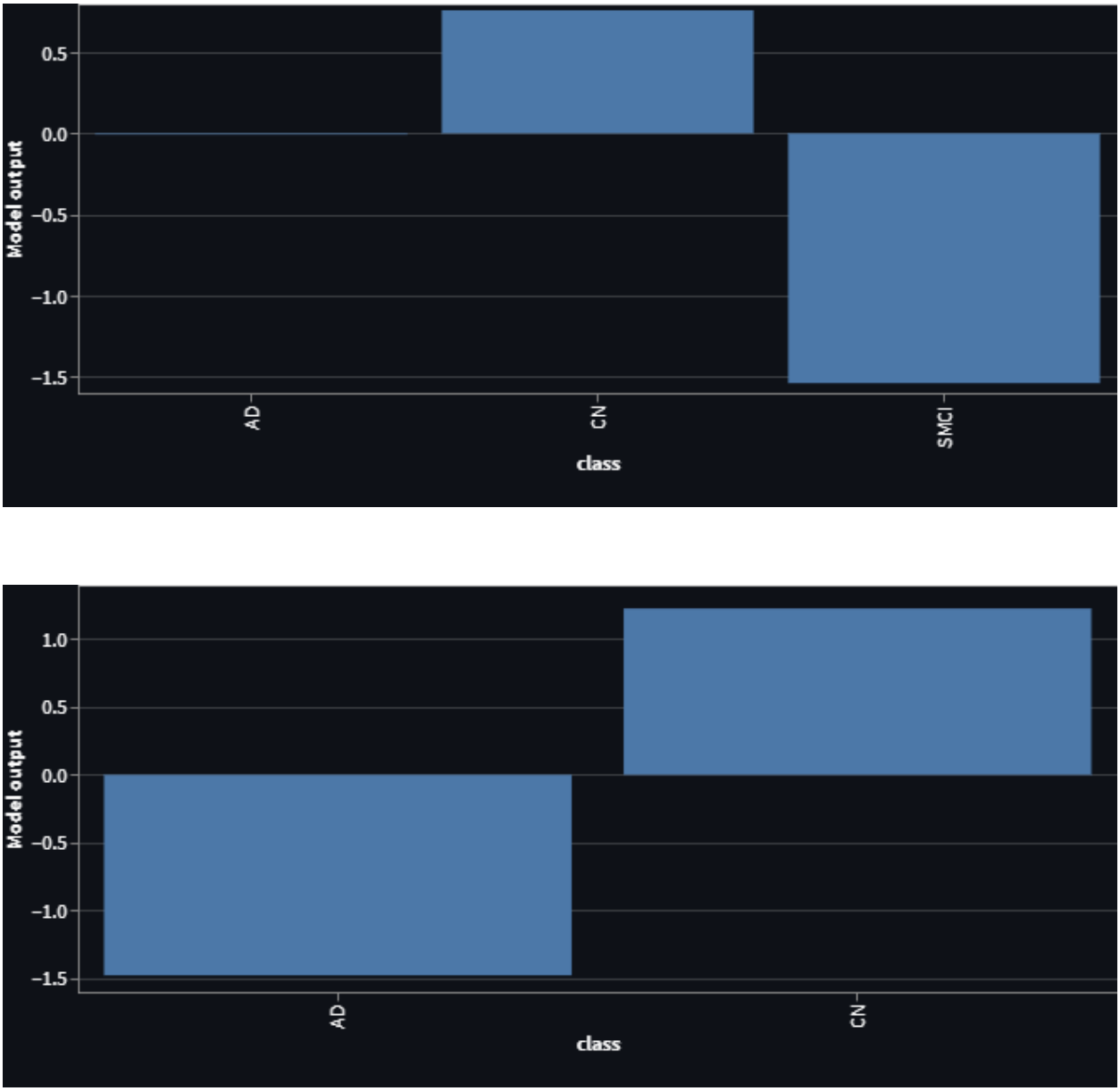


Figure 6-8 Case 2: Output of both the models

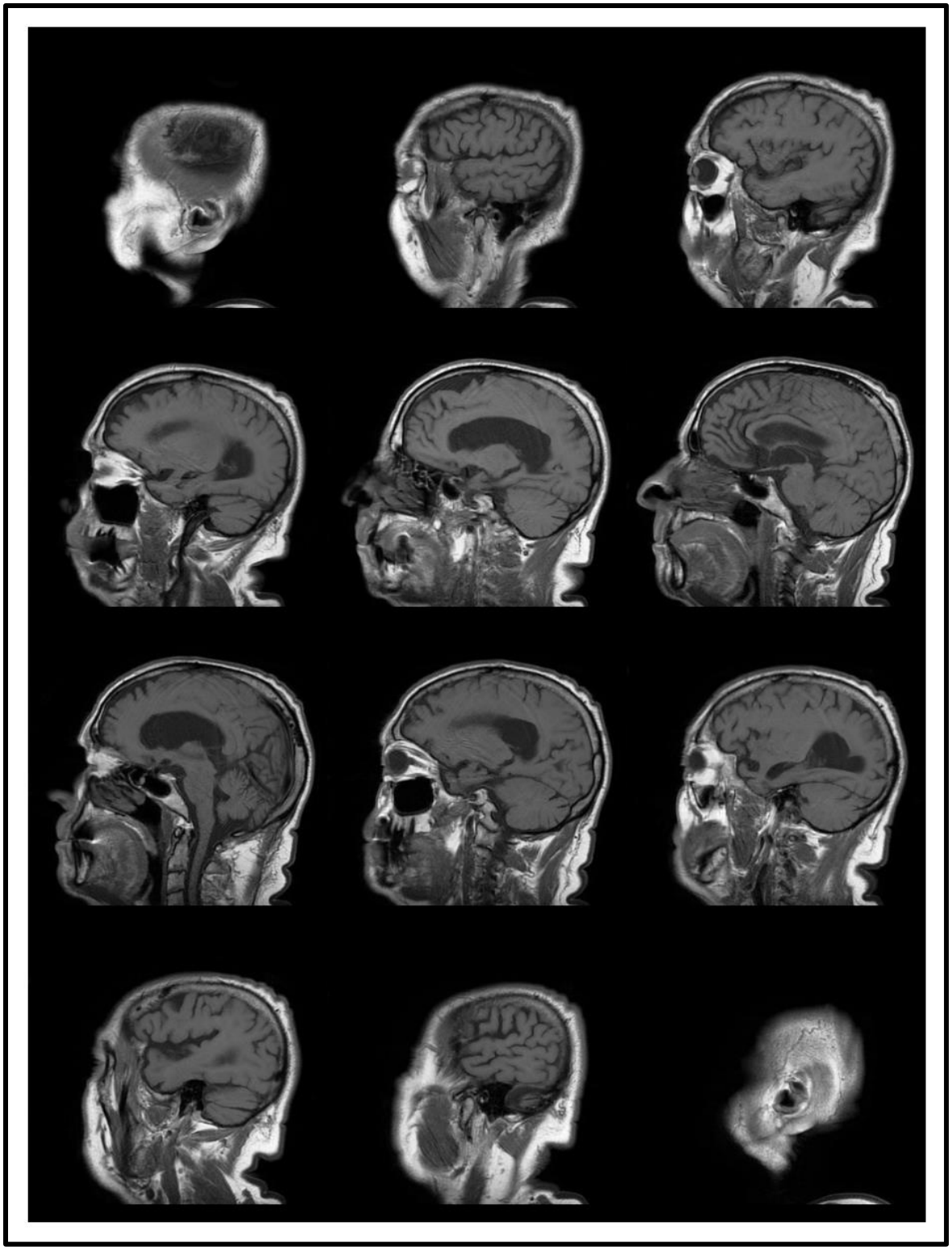


Figure 6-9 Case 3 MRI Scan Input

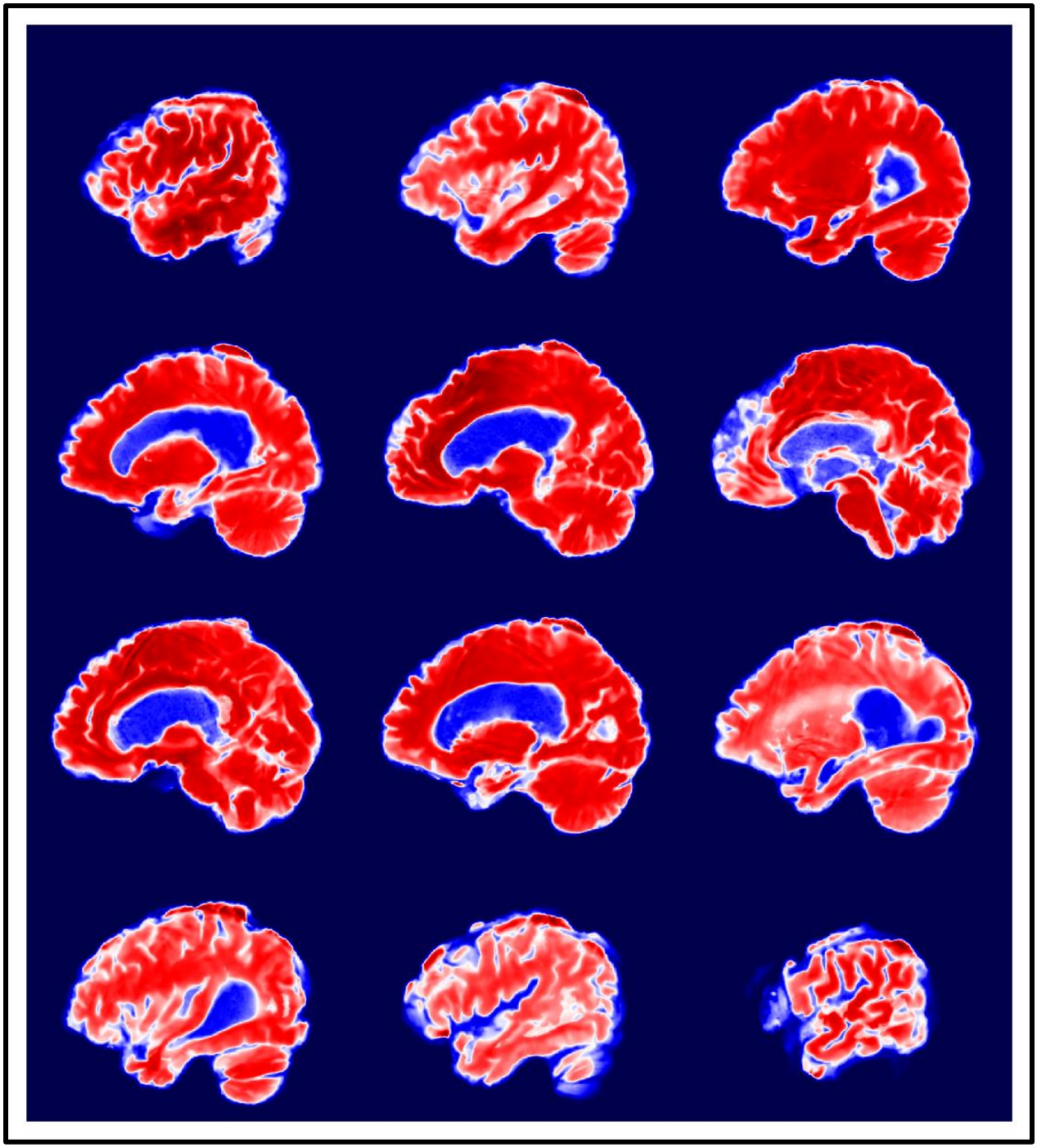


Figure 6-10 Case 3 MRI Scan After Preprocessing

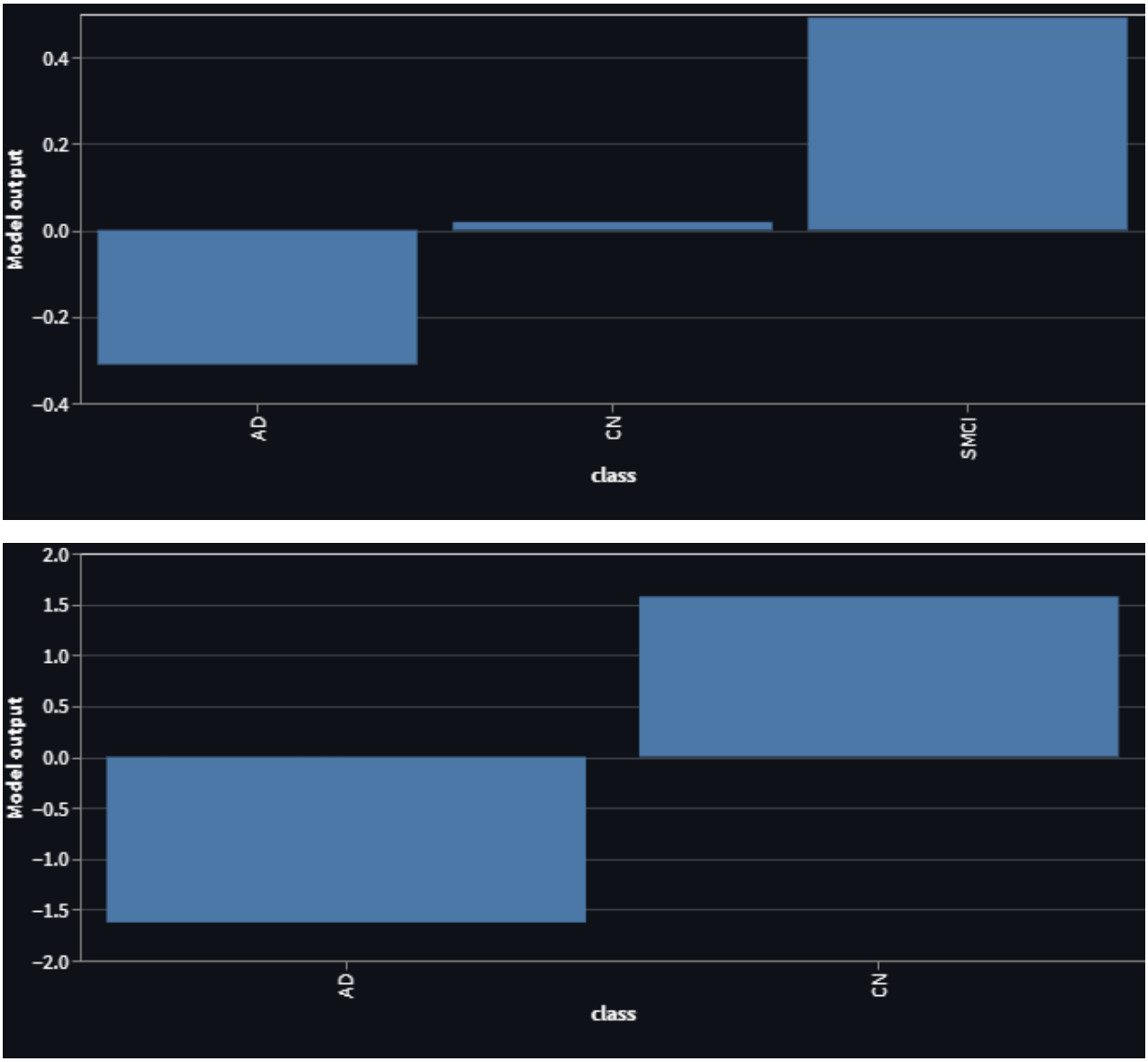


Figure 6-11 Case 3: Output of both the models

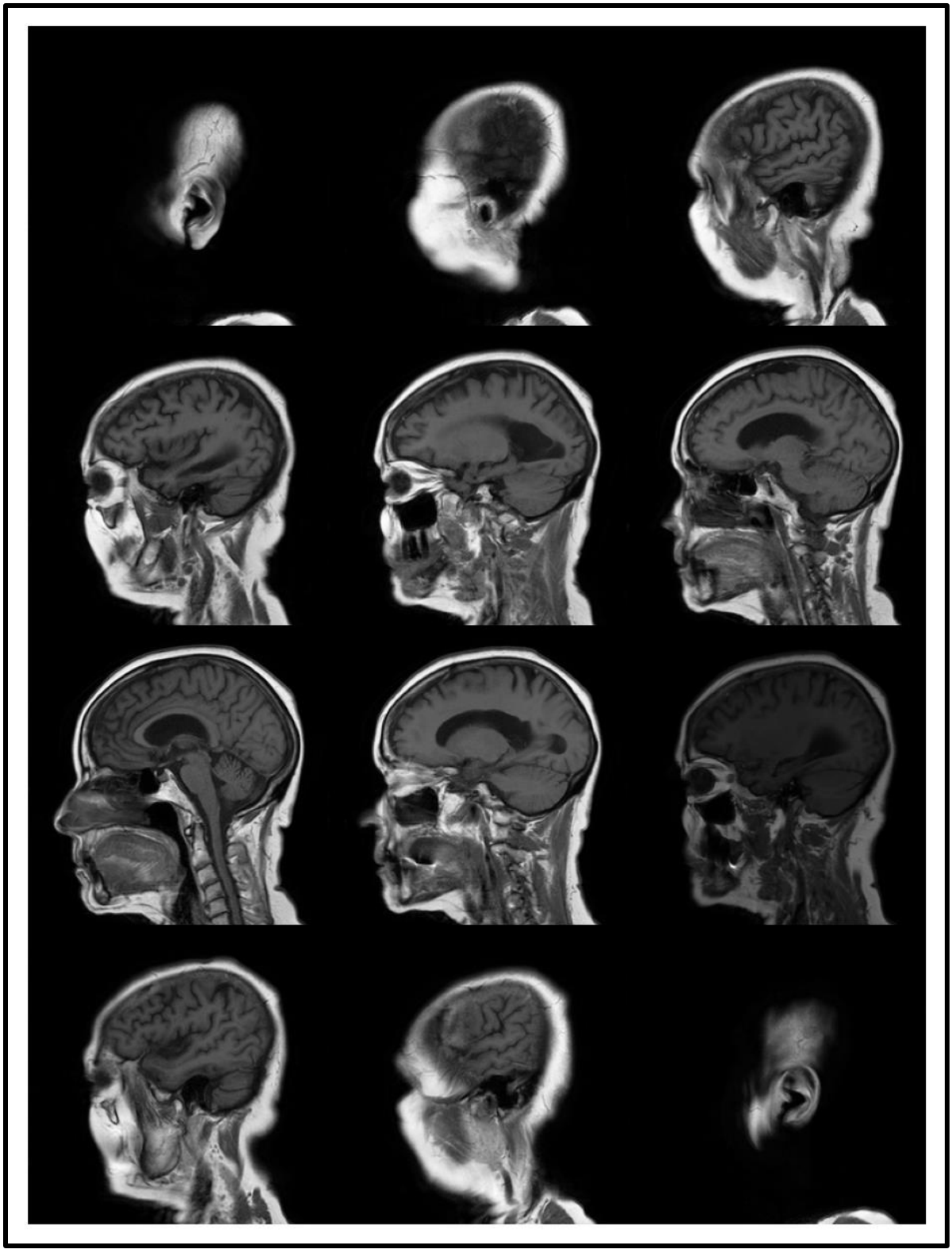


Figure 6-12 Case 4 MRI Scan Input

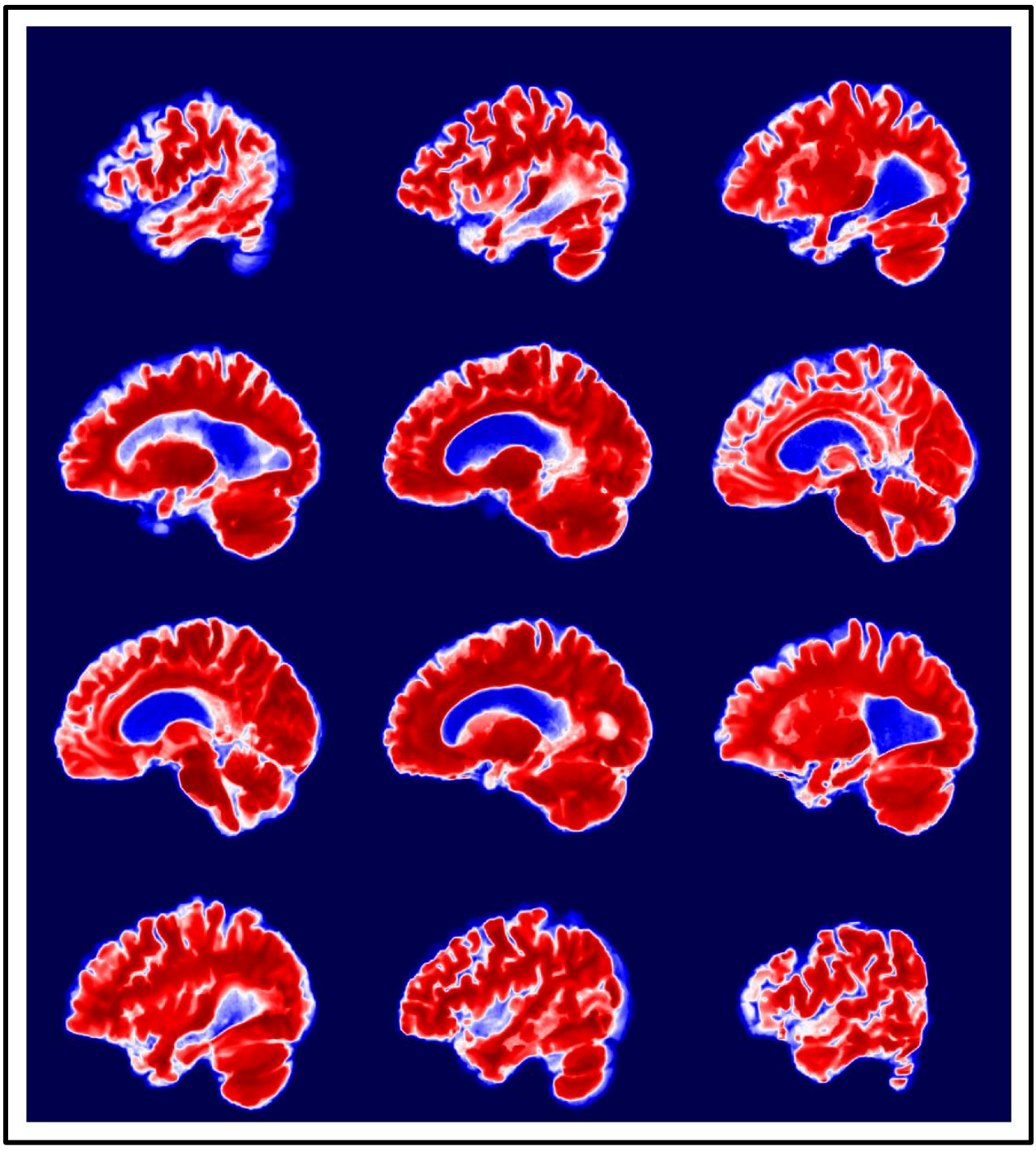


Figure 6-13 Case41 MRI Scan After Preprocessing

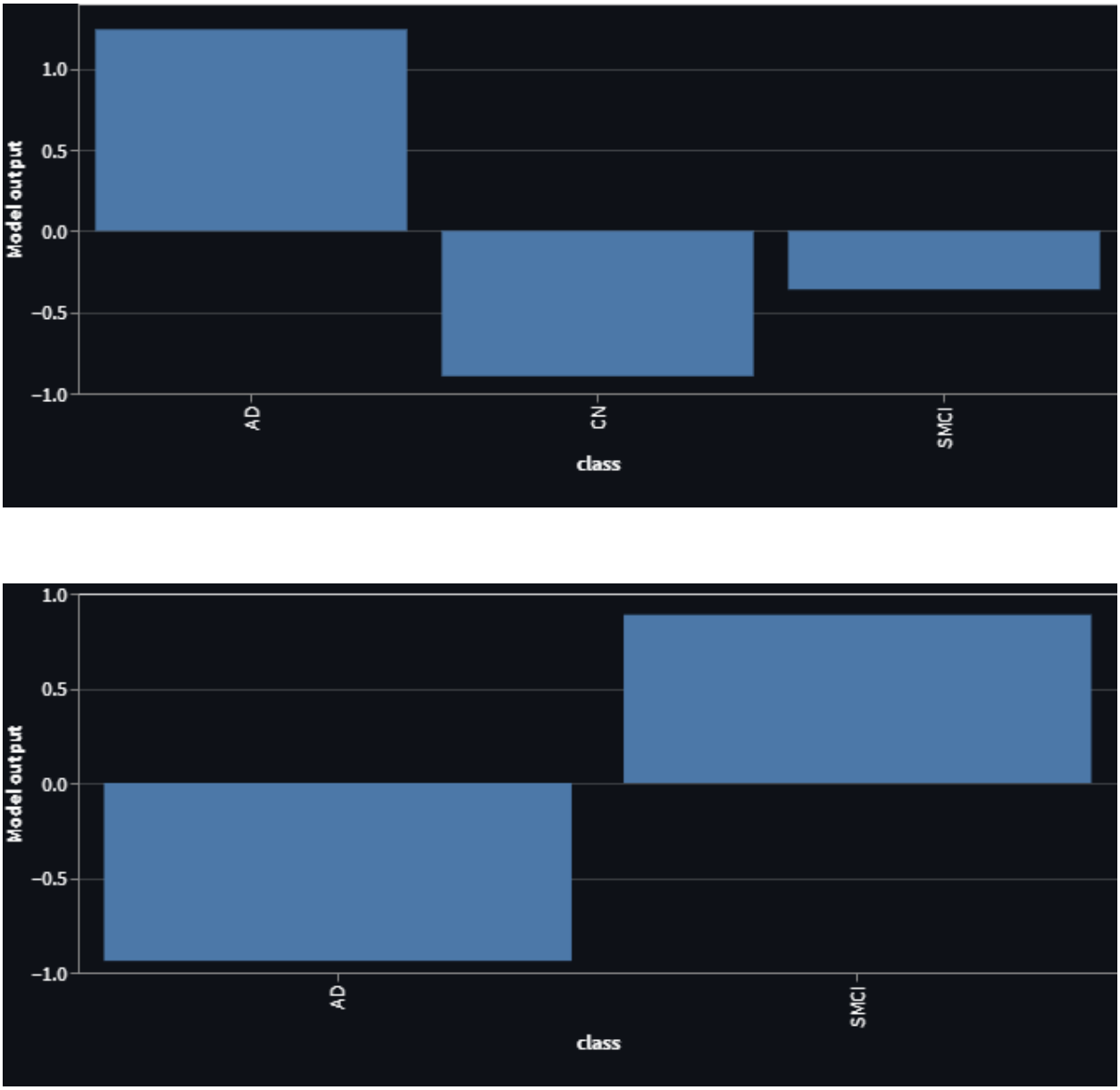


Figure 6-14 Case 4: Output of both the models

6.4 Conclusion

In the previous chapter DEEP-AD was proposed as a tool to help neuroradiologist in taking decision about the early detection of AD. In this chapter We **evaluated DEEP-AD by using Spanish dataset as a contribution toward attaining Goal #3.** It provided valuable insights into the performance of the application in a real-world setting.

The **results of the evaluation revealed that DEEP-AD had a high degree of accuracy**. The diagnostic **outcomes produced by the application were in concordance with the clinical diagnoses obtained from the neuroradiologists**.

This serves as solid evidence that the application has the potential to provide dependable diagnostic outcomes. Moreover, the neuroradiologists provided feedback on the application and found it to be user-friendly, intuitive, and easy to use. The **application provides clear diagnostic outcomes, and the results are easy to interpret.**

However, neuroradiologists suggest DEEP-AD requires more training to increase its accuracy and generalizability by utilizing real clinical data and identification of the factors of the MRI scans that seem to have played the most important part in the choice-making process.

Chapter 7: Conclusions and Future Work

Summary: This chapter offers the dissertation's overarching findings and conclusions, describing the primary research topic **(how to aid neuroradiologist in the early diagnosis of AD via the use of neuroimaging biomarkers**) and outlining the suggested aims and contributions. Study findings offered sufficient proof that these objectives were met. Based on the findings of these empirical investigations, we propose new avenues for research. Supporting the relevance, originality, and importance of this dissertation's goals and contributions are **two publications in JCR-SCI peer-reviewed journals**. Additionally, **one paper has been submitted to Springer Nature's Journal of Medical Systems** and is under the peer-review process. All these articles back up the findings presented in this dissertation as well.

7.1 Conclusions

Results from earlier research suggest that applying machine learning algorithms to the field of medical image analysis might have positive consequences (for example, helping doctors determine the severity of a certain condition by using biomarkers). However, using these tactics often necessitates the involvement of technical personnel and requires several iterations of optimization, both of which may be time-consuming and challenging.

AD is the most common type of dementia, and there is no known treatment for it currently. It is believed that at least 50 million people worldwide now suffer from AD and other dementias. The progression of AD occurs gradually over many years before clinical symptoms show. Studying innovative early detection techniques for various forms of dementia, such as AD or MCI, is essential to ensuring proper care and halting the disease's progression.

Given this context, the goal of this dissertation is to create the theoretical and technical tools by using DL needed to diagnose AD in its early stages more quickly and accurately, so that patients can get treatment right away.

Consequently, at the beginning of this dissertation, we explored the current body of research regarding applied DL in the early detection of AD. To this end, we conducted a systematic literature review, as reported in Chapter 2**.** This review of the literature was published in **[98]**.

With this context, to contribute with a series of empirical studies to understand the impact of DL-based tactics in real-world clinical settings and to acquire helpful insights for the creation of conceptual and technical tools, we formulated the first intermediate goal **(GOAL #1)** of this

dissertation: "To understand the DL models that have performed well at identifying patterns associated with AD by using the neuroimaging biomarkers." To this end, we carried out thirteen empirical studies aiming to understand and compare the effect of DL strategies on neuroimaging biomarkers (Cont. #1, Chapter 2)**.**

Results show that most of the research made use of CNNs with various TL methods. However, certain very well-known TL designs that have been very helpful in the analysis of medical data, like EfficientNet and DenseNet, have not been implemented for AD. Several issues need to be rectified, including overfitting in relation to the use of small datasets and reproducibility in respect to the randomness occurring during training. These studies have been published in [19], [22], [137]–[147].

Because of these limitations, it is likely that AI-based systems that help doctors make decisions about the current state of AD are not available. To assist in overcoming such limitations, we proposed the second intermediate goal (GOAL #2), which is to provide the neuroradiologist with the computer-interpretable information they need to analyze an MRI scan. This goal has been achieved by the following contributions (CONT#2)**: (1**) the identification and implementation of the most efficient DL model using E2EL (Chapter 3)**;** and (2) the improvement of the accuracy of the identified model through the utilization of the fusion of E2EL and TL (Chapter 4)**.**

Eight DL models based on 3D CNN from the DenseNet and EfficientNet families with E2EL have been implemented (Chapter 3). Following implementation, we conducted a comparative study utilizing an efficiency effects graph and a rank mechanism to choose the optimal model. The results of this analysis revealed that the EfficientNetB0 and DenseNet264 models performed the best among all models for binary classification tasks to categorize (AD, sMCI) and (AD, CN) subjects (Chapter 3). The work done in Chapter 3 has been published in [101].

To enhance the accuracy of the AD vs. sMCI task, we proposed and implemented a novel method termed "Fusion of E2EL and TL" to improve the accuracy of EfficienNet-B0 (Chapter 4). In addition, for implementing the ensemble learning-based tool, I built a DL model for the multiclass classification task to categorize (AD, sMCI, and CN) subjects (Chapter 4).

A research paper for Chapter 4 has been submitted and is currently being reviewed by the Springer Nature Journal of Medical Systems.

To verify model's responsiveness in real clinical settings, the deployed system must also be conveniently accessible via some online medium to assist neuroradiologists. With this context we proposed the final intermediate goal (GOAL#3) A tool that neuroradiologists may use at their convenience to evaluate the model's robustness.

Therefore, "DEEP-AD," a web-based tool to aid neuroradiologists in their clinical judgment regarding the early detection of AD, is another significant contribution (CONT#3) of this dissertation. DEEP-AD is founded on an ensemble of 3D CNNs (EfficientNet-B0 and DenseNet-264 architectures) with E2EL and 3DTL. Furthermore, it has been validated on 41 individuals using Spanish MRI datasets, with 82.92% success rate. This data collection was given by HT Medica, Spain.

To sum up, this dissertation tackled the issue of aiding neuroradiologists in making decisions related to early diagnosis of AD. Thirteen empirical studies were examined to gain a better understanding of the effectiveness of these methods and to obtain insights for the creation of DL models that could accurately detect early-stage AD. Additionally, a fully operational prototype of the proposed DEEP-AD system was developed, confirming the assertion of neuroradiologists that these methods can be designed, implemented, and managed within a reasonable budget.

7.2 Future Lines of Work

This doctoral dissertation's proposed contributions and goals have made it possible to find new ways to study the role of DL in the early detection of Alzheimer's disease. These ideas for future research have been put into two groups: (a) future research that could build on the results of the dissertation, and (b) ways that the dissertation's contributions could be used in new areas of applied DL research.

7.2.1 Research Extension

This subsection presents future research lines of work associated with the goals pursued in this dissertation:

- Suggested by neuroradiologist (Chapter 6): ‘ DEEP-AD needs further training using data from several hospitals to improve its accuracy and generalizability and find out the factors that seem to have had the most significant role in the decision-

making process'. The models used in DEEP-AD can be trained further by utilizing actual clinical data, as we have only utilized the data from ADNI and IXI during training. Additionally, various techniques, such as activation maps and filter visualization, may be added to identify the features of MRI scans used for making the decision.

- The EfficientNet-B0 used for the multiclassification task **(Chapter 4)** may be further enhanced by employing the fusion of E2EL and TL, which may enhance DEEP AD's overall prediction accuracy.
- Researchers may train their 3D models utilizing this novel method, called "Fusion of E2EL and TL" (Chapter 4), especially when they have minimal datasets and wish to analyze 3D data without splitting it into 2D slices or concentrating on ROI.
- The Spanish data set of 41 MRI scans of MCI, CN, and AD participants (Chapter 6) obtained from HT Medica, Spain, may be used by researchers in their study. Because it is unique and incorporates MRI scans from several hospitals in Spain, and it is not accessible online. Researchers who want to develop a generalizable model that is focused on the Spanish Regime may utilize this dataset.
- 3D Transfer Learning: Researchers may analyze their 3D data (not necessarily medical data) by using the models (EfficientNe-B0 and DenseNet264) that were trained and validated in chapter 3 as 3D pretrained models .
- The analysis of the thirteen empirical studies outlined in chapter 2 could be used to gain a better understanding of applied DL in the health domain or to analyze other types of 3D data.

7.2.2 Potential Applications

Some of the work done for this dissertation has prompted the suggestion of possible DL uses in medical image analysis. These are emergent research lines, as opposed to the ones that came before them; hence, they are not immediately connected to the dissertation's goals.

- **Multimodality** approach performs better then single modality. In the future, new combinations of biomarkers may be used to predict the onset of AD.

- The dearth of substantial neuroimaging datasets causes generalizability issues, despite the use of TL and augmentation. To make a generalizable classifier, large datasets can be created using neuroimaging data synthesis.
- It may be challenging to show precisely which features have been retrieved and to regulate how those features affect the inference and relative prominence of other characteristics. As a result, it is challenging to eliminate any biases that the input datasets may have generated.
- Even with the endemic issues of neuroimaging, where training data are few and sample dimensionality is high, DL models with E2EL and TL can be used to obtain good accuracy, However, it requires the fine tuning of hyperparameters and an appropriate 3D CNN architecture specifically designed for TL with excellent potential for generalization; additionally, MRI scans must be thoroughly pre-processed to maintain the spatial link and enhance image quality.
- To initialize the weights of 3D models, 3D pretrained models may be created, realizing that we couldn't identify a CNN architecture that supports 3D TL.
- We also observed that MONAI offers a leading framework to implement DL models for medical image analysis, as it is simple to understand and supports a wide range of functions. Additionally, Google Colab Pro+ offered the best online cloud-based resources, with access to large RAM and excellent GPUs, thus enabling us to achieve this task despite certain drawbacks such as GPU unavailability under heavy load.

7.3 Publications

The number of publications that reference the research discussed throughout this dissertation demonstrates the importance and effectiveness of this research.

This section lists the different published documents describing part of the work and the results obtained from this research process. The list only includes peer-reviewed publications in which the dissertation author is first author.

Publications in JCR-indexed international journals:

J1 **[JCR-SCI Q2]** : Agarwal, D.; Marques, G.; de la Torre-Díez, I.; Franco Martin, M.A.; García Zapiraín, B.; Martín Rodríguez, F. **Transfer Learning for Alzheimer's Disease through Neuroimaging Biomarkers: A Systematic Review.** *Sensors* **2021**, *21*, 7259. https://doi.org/10.3390/s21217259

J2 **[JCR-SCI Q1] :** Agarwal, D.; Berbis, M.A.; Martín-Noguerol, T.; Luna, A.; Garcia, S.C.P.; de la Torre-Díez, I. **End-to-End Deep Learning Architectures Using 3D Neuroimaging Biomarkers for Early Alzheimer's Diagnosis**. *Mathematics* **2022**, *10*, 2575. https://doi.org/10.3390/math10152575

Submitted in JCR-indexed international journal: Under peer review

J3 **[JCR-SCI Q1] : Automated medical diagnosis of Alzheimer´s Disease using an Efficient Net Convolutional Neural Network** Springer Nature, Journal of Medical Systems 2023 ,Submission id : ' 7d42cab4-7701-4c83-8a1f-2ad4a775b640'

[Accepted and published on 2/05/2023 : https://link.springer.com/article/10.1007/s10916-023-01941-4]

Appendix A :Technical insights into the included articles in SLR

The technical aspects of thirteen empirical studies identified and analyzed during SLR (**reported in Chapter 2**) are provided in this appendix.

A.1 DL models, TL techniques and validation procedure

The following table shows the all the specifics of the DL models used in the thirteen empirical studies.

Ref	Dataset	Sample size	DL Arch.	TL Method	Modality	Input	Validation
[137]	ADNI	CN-25, SMC-25, EMCI-25, LMCI-25, MCI-13, AD-25	2D CNN	ResNet-18	fMRI	2D Slices	Random Sampling
[140]	ADNI	AD-294, ncMCI-510, cMCI-253, CN-352	3D CNN	Local TL	T1W MRI	Pre-Processed 3D Scans	10-fold CV
	MILAN (Local)	AD-124, ncMCI-23, cMCI-27, CN-55					
[141]	ADNI	AD-198, CN-230, pMCI-166, sMCI-101	3D CNN	CAE & Local TL	T1W MRI	Pre-Processed 3D Scans	5-fold CV
[142]	ADNI	CN-330, MCI-787, sMCI-298, pMCI-295, AD-336	3D & 2D CNN	CAE & ResNet-18	T1W MRI	Pre-Processed 3D Scans & 2D Slices	5-fold CV
	OASIS	CN-76, AD-78					
	AIBL	CN-429, MCI-93, sMCI-13, pMCI-20, AD-76					
[22]	ADNI	CN -237, sMCI -245, pMCI -189, AD -157	3D CNN	Local TL	T1W MRI	Segmented scans [GM Tissues]	5-fold CV
[19]	ADNI	CN-150, sMCI-150, cMCI-157	2D CNN	GoogleNet & CaffeNet	T1W MRI	2D Slices	5-fold CV
[143]	ADNI	cMCI-61, ncMCI-276	2D CNN	CaffeNet	T1W MRI	Segmented Scans [WM,GM,CSF] Converted into 2D PNG Images	10-fold CV
[144]	ADNI	CN-304, sMCI-409, pMCI-112, AD-226	DNN	SAE & Local TL	FDG-PET + T1WMRI	Segmented scans [WM/GM Patches]	10-fold CV

Ref	Dataset	Sample size	DL Arch.	TL Method	Modality	Input	Validation
[145]	ADNI	CN-182, AD-139, cMCI-79, ncMCI-92	3D CNN	Local TL	18F-FDG & AV-45 PET Scans	PET Scans	10-fold CV
[146]	ADNI	cMCI-70, ncMCI-280	2D CNN	AlexNet	18F-FDG PET	2D Slices	No
[147]	OASIS	100-AD, CN-100	2D CNN	VGG 16 & Inception v3	T1W MRI	2D Slices	5-fold CV
[138]	ADNI	AD-194, pMCI-164, sMCI-233, CN-216	DenseNets & RNN	Local TL	T1W MRI	Segmented Scans [Hippocampi patches]	5-fold CV
[139]	ADNI	CN-256.cMCI-168, cMCI-129	3D CNN	Local TL	T1W MRI	Pre-Processed 3D Scans	5-fold CV
	IXI	581-NC					

A.2 Preprocessing pipeline

The table below shows how different pre-processing strategies for neuroimaging biomarkers are used by category, as well as the software tool used for pre-processing.

Ref.	Pre-processing-Pipeline	Category	Software
[137]	Brain extraction, Motion correction, Slice timing correction Intensity normalization, Spatial smoothing- Gaussian Kernel (FWHM), High-pass filtering, Spatial normalization.	Slice-based	FSL
[140]	Registration in the MNI space, Segmentation to yield GM, WM, and CSF tissue probability maps.	Voxel-based	SPM 12
[141]	Bias-field correction, Registration in the MNI space.	Voxel-based	SPM 12
[142]	Bias field correction (BY N4ITK), Registration (Affine BY SyN) in MNI space, Cropping, Intensity rescaling.	Voxel-based	Nipype
[22]	Segmentation in GM tissues, Spatial normalization, Spatial smoothing [3D Gaussian kernel to 6 mm FWHM].	ROI-Based	SPM12
[19]	Gradwarp, Intensity inhomogeneity correction, N3 histogram peak sharpening.	Slice-based	MATLAB
[143]	Segmentation into GM. WM and CSF tissues, GM maps were modulated using the Jacobean determinants, Smoothing using an 8-mm FWHM Gaussian kernel, scaling with trilinear interpolation, Images were exported in a PNG format, PCA and sequential feature selection.	Voxel-based	SPM12
[144]	MRI segmented into GM and WM tissues, GM and WM were further subdivided into patches of varying sizes, Co-registration of FDG-PET image, and segmented MRI scans.	Patch-Based	FreeSurfer 5.3 package
[145]	Already pre-processed PET images were downloaded, co-registration of FDG and AV-45 PET images,	Voxel-based	---
[146]	Interpolation, Segmentation of each scan from the Z-axis into 65 2D Slices.	Slice-based	NifTi_2014 toolkit
[147]	Choose the most explanatory 32 2D slices from the axial plane by sorting entropies.	Slice-based	MATLAB
[138]	Affine registration, Segmentation of the hippocampus, extraction of a fixed size 3D patch from the center of each hippocampus.	Patch-Based	FSL
[139]	Rigid registration.	Voxel-based	MRIcron

Appendix B :Links for accessing the preprocessed datasets and scripts

- Access the scripts that have been built for preprocessing MRI scans and creating the 5 datasets for performing stratified 5-fold CV.

  https://colab.research.google.com/drive/14NwikUAnM10mko3jqlrCzoZD9VxYeOtb?usp=sharing

- Access preprocessed datasets of T1W MRI scans acquired from ADNI and IXI, utilized in chapters 3 and 4.
  - For accessing Dataset for Fold -1 https://drive.google.com/drive/folders/1UmcuYryjTvzYlzDtuVnssaGK60ppH9T?usp=sharing
  - For accessing Dataset for Fold -2 https://drive.google.com/drive/folders/1cikT3_BENynyTlfCxPIflMiPIBsQsZ6?usp=sharing
  - For accessing Dataset for Fold -3 https://drive.google.com/drive/folders/14PxZk2m_6kFSeKDM9BLprJvPgYiaBs4Q?usp=sharing
  - For accessing Dataset for Fold -4 https://drive.google.com/drive/folders/19Yh3hb5Iv4adlBmmgqFgHdfEnUX8N600?usp=sharing
  - For accessing Dataset for Fold -5 https://drive.google.com/drive/folders/1FfAoY41DGs4TBIlauDFNund2jJhDPX8L?usp=sharing
- Access the scripts of Training/Validation & Testing of eight models (DenseNet-121, 169, 201, 264) and (EfficientNet-B0, B1, B2, B3) as reported in Chapter 3 for binary classification tasks AD vs. CN and AD vs. sMCI by using E2EL.
  - For AD vs. CN task https://drive.google.com/drive/folders/1tLBDmYoi97ORnKinE-i0nIl-S8X-LzQf?usp=sharing

    - For AD vs. sMCI task https://drive.google.com/drive/folders/1KmjtHllXl_FqFD_ddX3ByHYwMgYRAuUB?usp=share_link
- Access the scripts of Training/Validation & Testing of EfficientNet-B0 as **reported in Chapter 4** for binary classification task of AD vs. sMCI by using the fusion of 'E2EL and TL'.
    - https://colab.research.google.com/drive/13Uqvn6fnMsmjAVF5qMbdGi-gdE1GymAR?usp=sharing
- Access the scripts of Training/Validation & Testing of EfficientNet-B0 as **reported in Chapter 4** for multiclass classification task of AD vs. CN vs. sMCI by using the E2EL.
    - https://drive.google.com/drive/folders/1d-FCa3gWOS48YXEHw-LcKLP6Hw1K9T1K?usp=share_link

Appendix C : DEEP-AD Validation

C.1 Access the MRI scans of Spanish datasets

Researchers can access the Spanish datasets of MRI scans validated by DEEP-AD for this research. [can be shared upon reasonable request]

C.2 Collaboration and Validation/Feedback Report Given by neuroradiologist

The validation report and feedback given by **Dr. M. Alvaro Berbís** Head of R&D and Innovation HT Médica Madrid, Spain.

- Collaboration certificate with HT Médica Madrid, Spain
    - https://drive.google.com/file/d/1NLe3NfMJ-mmk1ZjpKHL-_tnrUKmnTBC6/view?usp=sharing
- Validation Report & Feedback link
    - https://drive.google.com/file/d/1QYlsBnP5dNArcyZpYbKl0Jcr9_3PX4m-/view?usp=sharing

C.3 Screen shots of the prediction of all MRI scans of Spanish datasets through DEEP-AD

https://drive.google.com/drive/folders/1eG2foKJUTjOztb9p1kYiZgS3O1m6D8qG?usp=sharing

Bibliography

[1] J. Hardy, “Amyloid, the presenilins and Alzheimer’s disease,” *Trends Neurosci.*, vol. 20, no. 4, pp. 154–159, May 1997, doi: 10.1016/S0166-2236(96)01030-2.

[2] “Alzheimer’s Disease Facts and Figures,” *Alzheimer’s Disease and Dementia*. https://www.alz.org/alzheimers-dementia/facts-figures (accessed Apr. 29, 2022).

[3] G. A. Jicha *et al.*, “Neuropathologic Outcome of Mild Cognitive Impairment Following Progression to Clinical Dementia,” *Arch. Neurol.*, vol. 63, no. 5, pp. 674–681, May 2006, doi: 10.1001/archneur.63.5.674.

[4] “Big Data Challenge for Alzheimer’s Launches in Global Effort to Use Innovative Open Science Techniques to Improve Diagnosis and Treatment,” *UsAgainstAlzheimer’s*. https://www.usagainstalzheimers.org/press/big-data-challenge-alzheimer%E2%80%99s-launches-global-effort-use-innovative-open-science-techniques (accessed Dec. 16, 2022).

[5] “A Machine learning neuroimaging challenge for automated diagnosis of Mild Cognitive Impairment,” *ResearchGate*. https://www.researchgate.net/project/A-Machine-learning-neuroimaging-challenge-for-automated-diagnosis-of-Mild-Cognitive-Impairment (accessed Dec. 16, 2022).

[6] R. Cuingnet *et al.*, “Automatic classification of patients with Alzheimer’s disease from structural MRI: A comparison of ten methods using the ADNI database,” *NeuroImage*, vol. 56, no. 2, pp. 766–781, May 2011, doi: 10.1016/j.neuroimage.2010.06.013.

[7] A. Retico, P. Bosco, P. Cerello, E. Fiorina, A. Chincarini, and M. E. Fantacci, “Predictive Models Based on Support Vector Machines: Whole-Brain versus Regional Analysis of Structural MRI in the Alzheimer’s Disease,” *J. Neuroimaging*, vol. 25, no. 4, pp. 552–563, 2015, doi: 10.1111/jon.12163.

[8] Y. Zhu, X. Zhu, M. Kim, D. Shen, and G. Wu, “Early Diagnosis of Alzheimer’s Disease by Joint Feature Selection and Classification on Temporally Structured Support Vector Machine,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, Cham, 2016, pp. 264–272. doi: 10.1007/978-3-319-46720-7_31.

[9] I. Beheshti, H. Demirel, and H. Matsuda, “Classification of Alzheimer’s disease and prediction of mild cognitive impairment-to-Alzheimer’s conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm,” *Comput. Biol. Med.*, vol. 83, pp. 109–119, Apr. 2017, doi: 10.1016/j.compbiomed.2017.02.011.

[10] Z. Sun, Y. Qiao, B. P. F. Lelieveldt, and M. Staring, “Integrating spatial-anatomical regularization and structure sparsity into SVM: Improving interpretation of Alzheimer’s disease classification,” *NeuroImage*, vol. 178, pp. 445–460, Sep. 2018, doi: 10.1016/j.neuroimage.2018.05.051.

[11] N. Zeng, H. Qiu, Z. Wang, W. Liu, H. Zhang, and Y. Li, “A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer’s disease,” *Neurocomputing*, vol. 320, pp. 195–202, Dec. 2018, doi: 10.1016/j.neucom.2018.09.001.

[12] S. Rathore, M. Habes, M. A. Iftikhar, A. Shacklett, and C. Davatzikos, “A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer’s disease and its prodromal stages,” *NeuroImage*, vol. 155, pp. 530–548, Jul. 2017, doi: 10.1016/j.neuroimage.2017.03.057.

[13] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[14] J. Samper-González *et al.*, “Reproducible evaluation of classification methods in Alzheimer’s disease: Framework and application to MRI and PET data,” *NeuroImage*, vol. 183, pp. 504–521, Dec. 2018, doi: 10.1016/j.neuroimage.2018.08.042.

[15] J. Zhang, B. Zheng, A. Gao, X. Feng, D. Liang, and X. Long, "A 3D densely connected convolution neural network with connection-wise attention mechanism for Alzheimer's disease classification," *Magn. Reson. Imaging*, vol. 78, pp. 119–126, May 2021, doi: 10.1016/j.mri.2021.02.001.
[16] A. Mehmood, M. Maqsood, M. Bashir, and Y. Shuyuan, "A Deep Siamese Convolution Neural Network for Multi-Class Classification of Alzheimer Disease," *Brain Sci.*, vol. 10, no. 2, p. 84, Feb. 2020, doi: 10.3390/brainsci10020084.
[17] B. Solano-Rojas and R. Villalón-Fonseca, "A Low-Cost Three-Dimensional DenseNet Neural Network for Alzheimer's Disease Early Discovery," *Sensors*, vol. 21, no. 4, p. 1302, Feb. 2021, doi: 10.3390/s21041302.
[18] M. Odusami, R. Maskeliūnas, R. Damaševičius, and T. Krilavičius, "Analysis of Features of Alzheimer's Disease: Detection of Early Stage from Functional Brain Changes in Magnetic Resonance Images Using a Finetuned ResNet18 Network," *Diagn. Basel Switz.*, vol. 11, no. 6, p. 1071, Jun. 2021, doi: 10.3390/diagnostics11061071.
[19] C. Wu *et al.*, "Discrimination and conversion prediction of mild cognitive impairment using convolutional neural networks," *Quant. Imaging Med. Surg.*, vol. 8, no. 10, pp. 992003–991003, Nov. 2018.
[20] A. A, P. M, M. Hamdi, S. Bourouis, K. Rastislav, and F. Mohmed, "Evaluation of Neuro Images for the Diagnosis of Alzheimer's Disease Using Deep Learning Neural Network," *Front. Public Health*, vol. 10, 2022, Accessed: Apr. 25, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpubh.2022.834032
[21] E. Goceri, "Diagnosis of Alzheimer's disease with Sobolev gradient-based optimization and 3D convolutional neural network," *Int. J. Numer. Methods Biomed. Eng.*, vol. 35, no. 7, p. e3225, 2019, doi: 10.1002/cnm.3225.
[22] A. Abrol, M. Bhattarai, A. Fedorov, Y. Du, S. Plis, and V. Calhoun, "Deep residual learning for neuroimaging: An application to predict progression to Alzheimer's disease," *J. Neurosci. Methods*, vol. 339, p. 108701, Jun. 2020, doi: 10.1016/j.jneumeth.2020.108701.
[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, 2012, vol. 25. Accessed: Oct. 11, 2022. [Online]. Available: https://papers.nips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html
[24] G. Marques, D. Agarwal, and I. de la Torre Díez, "Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network," *Appl. Soft Comput.*, vol. 96, p. 106691, Nov. 2020, doi: 10.1016/j.asoc.2020.106691.
[25] S. N. Wadekar *et al.*, "Towards End-to-End Deep Learning for Autonomous Racing: On Data Collection and a Unified Architecture for Steering and Throttle Prediction," *ArXiv210501799 Cs*, May 2021, Accessed: Apr. 30, 2022. [Online]. Available: http://arxiv.org/abs/2105.01799
[26] G. Mesnil *et al.*, "Unsupervised and Transfer Learning Challenge: a Deep Learning Approach," in *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, Jun. 2012, pp. 97–110. Accessed: Oct. 10, 2022. [Online]. Available: https://proceedings.mlr.press/v27/mesnil12a.html
[27] R. Raina, A. Y. Ng, and D. Koller, "Constructing informative priors using transfer learning," in *Proceedings of the 23rd international conference on Machine learning*, New York, NY, USA, Jun. 2006, pp. 713–720. doi: 10.1145/1143844.1143934.
[28] G. Marcus, "Deep Learning: A Critical Appraisal." arXiv, Jan. 02, 2018. doi: 10.48550/arXiv.1801.00631.
[29] "ADNI | Alzheimer's Disease Neuroimaging Initiative." https://adni.loni.usc.edu/ (accessed May 10, 2022).

[30] V. Gulshan *et al.*, "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, Dec. 2016, doi: 10.1001/jama.2016.17216.
[31] W. S. Bush and J. H. Moore, "Chapter 11: Genome-Wide Association Studies," *PLOS Comput. Biol.*, vol. 8, no. 12, p. e1002822, Dec. 2012, doi: 10.1371/journal.pcbi.1002822.
[32] K. D. Bailey, *Methods of social research*. New York : Free Press ; Toronto : Maxwell Macmillan Canada ; New York : Maxwell Macmillan International, 1994. Accessed: Dec. 22, 2022. [Online]. Available: http://archive.org/details/methodsofsocialr0000bail_e8m0
[33] "*The Structure of Scientific Revolutions*," *Wikipedia*. Sep. 24, 2022. Accessed: Dec. 22, 2022. [Online]. Available: https://en.wikipedia.org/w/index.php?title=The_Structure_of_Scientific_Revolutions&oldid=1112043223
[34] J. W. Creswell and C. N. Poth, *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. SAGE Publications, 2016.
[35] D. M. Mertens, *Research and evaluation in education and psychology: integrating diversity with quantitative, qualitative, and mixed methods*, 3rd ed. Los Angeles: SAGE Publications, 2010.
[36] C. Wohlin, M. Höst, and K. Henningsson, "Empirical Research Methods in Software Engineering," in *Empirical Methods and Studies in Software Engineering: Experiences from ESERNET*, R. Conradi and A. I. Wang, Eds. Berlin, Heidelberg: Springer, 2003, pp. 7–23. doi: 10.1007/978-3-540-45143-3_2.
[37] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A Design Science Research Methodology for Information Systems Research," *J. Manag. Inf. Syst.*, vol. 24, no. 3, pp. 45–77, Dec. 2007, doi: 10.2753/MIS0742-1222240302.
[38] J. F. Nunamaker, M. Chen, and T. D. M. Purdin, "Systems Development in Information Systems Research," *J. Manag. Inf. Syst.*, vol. 7, no. 3, pp. 89–106, Dec. 1990, doi: 10.1080/07421222.1990.11517898.
[39] S. Keele, "Guidelines for performing systematic literature reviews in software engineering," Technical report, ver. 2.3 ebse technical report. ebse, 2007.
[40] S. Grueso and R. Viejo-Sobera, "Machine learning methods for predicting progression from mild cognitive impairment to Alzheimer's disease dementia: a systematic review," *Alzheimers Res. Ther.*, vol. 13, no. 1, p. 162, Sep. 2021, doi: 10.1186/s13195-021-00900-w.
[41] T. Jo, K. Nho, and A. J. Saykin, "Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data," *Front. Aging Neurosci.*, vol. 11, 2019, Accessed: Apr. 24, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnagi.2019.00220
[42] R. Li, X. Wang, K. Lawler, S. Garg, Q. Bai, and J. Alty, "Applications of artificial intelligence to aid early detection of dementia: A scoping review on current capabilities and future directions," *J. Biomed. Inform.*, vol. 127, p. 104030, Mar. 2022, doi: 10.1016/j.jbi.2022.104030.
[43] J. M. Fernández Montenegro, B. Villarini, A. Angelopoulou, E. Kapetanios, J. Garcia-Rodriguez, and V. Argyriou, "A Survey of Alzheimer's Disease Early Diagnosis Methods for Cognitive Assessment," *Sensors*, vol. 20, no. 24, p. 7292, Dec. 2020, doi: 10.3390/s20247292.
[44] Y.-D. Zhang, S. C. Satapathy, X. Zhang, and S.-H. Wang, "COVID-19 Diagnosis via DenseNet and Optimization of Transfer Learning Setting," *Cogn. Comput.*, Jan. 2021, doi: 10.1007/s12559-020-09776-8.
[45] A. Shamila Ebenezer, S. Deepa Kanmani, M. Sivakumar, and S. Jeba Priya, "Effect of image transformation on EfficientNet model for COVID-19 CT image classification," *Mater. Today Proc.*, vol. 51, pp. 2512–2519, Jan. 2022, doi: 10.1016/j.matpr.2021.12.121.
[46] Y. Yang *et al.*, "A comparative analysis of eleven neural networks architectures for small datasets of lung images of COVID-19 patients toward improved clinical decisions," *Comput. Biol. Med.*, vol. 139, p. 104887, Dec. 2021, doi: 10.1016/j.compbiomed.2021.104887.

[47] I. Tarnanas, M. Tsolaki, T. Nef, R. M. Müri, and U. P. Mosimann, “Can a novel computerized cognitive screening test provide additional information for early detection of Alzheimer’s disease?,” *Alzheimers Dement.*, vol. 10, no. 6, pp. 790–798, 2014, doi: 10.1016/j.jalz.2014.01.002.
[48] Y. Abe, M. Toya, and M. Inoue, “Early detection system of senile dementia by behavior sensing: 2013 IEEE 17th International Symposium on Consumer Electronics, ISCE 2013,” *2013 IEEE 17th Int. Symp. Consum. Electron. ISCE 2013*, pp. 67–68, Sep. 2013, doi: 10.1109/ISCE.2013.6570255.
[49] S. Xefteris *et al.*, “Early Detection of Dementia: Advances, Challenges, and Future Prospects,” *Improving the Quality of Life for Dementia Patients through Progressive Detection, Treatment, and Care*, 2017. https://www.igi-global.com/chapter/early-detection-of-dementia/www.igi-global.com/chapter/early-detection-of-dementia/168927 (accessed Jan. 07, 2023).
[50] M. A. Ebrahimighahnavieh, S. Luo, and R. Chiong, “Deep learning to detect Alzheimer’s disease from neuroimaging: A systematic literature review,” *Comput. Methods Programs Biomed.*, vol. 187, p. 105242, Apr. 2020, doi: 10.1016/j.cmpb.2019.105242.
[51] T. Lewis, C. Synowiec, G. Lagomarsino, and J. Schweitzer, “E-health in low- and middle-income countries: findings from the Center for Health Market Innovations,” *Bull. World Health Organ.*, vol. 90, no. 5, pp. 332–340, May 2012, doi: 10.2471/BLT.11.099820.
[52] J. J. P. C. Rodrigues *et al.*, “Enabling Technologies for the Internet of Health Things,” *IEEE Access*, vol. 6, pp. 13129–13141, 2018, doi: 10.1109/ACCESS.2017.2789329.
[53] S. Sendra, L. Parra, J. Lloret, and J. Tomás, “Smart system for children’s chronic illness monitoring,” *Inf. Fusion*, vol. 40, pp. 76–86, Mar. 2018, doi: 10.1016/j.inffus.2017.06.002.
[54] M. J. Prince, A. Wimo, M. M. Guerchet, G. C. Ali, Y.-T. Wu, and M. Prina, “World Alzheimer Report 2015 - The Global Impact of Dementia: An analysis of prevalence, incidence, cost and trends,” Aug. 2015, Accessed: Jan. 07, 2023. [Online]. Available: https://kclpure.kcl.ac.uk/portal/en/publications/world-alzheimer-report-2015-the-global-impact-of-dementia(ae525fda-1938-4892-8daa-a2222a672254)/export.html
[55] G. Pandey and V. Ramakrishnan, “Invasive and non-invasive therapies for Alzheimer’s disease and other amyloidosis,” *Biophys. Rev.*, vol. 12, no. 5, pp. 1175–1186, Sep. 2020, doi: 10.1007/s12551-020-00752-y.
[56] J. M. Loomis, J. J. Blascovich, and A. C. Beall, “Immersive virtual environment technology as a basic research tool in psychology,” *Behav. Res. Methods Instrum. Comput.*, vol. 31, no. 4, pp. 557–564, Dec. 1999, doi: 10.3758/BF03200735.
[57] Z. Campbell, K. K. Zakzanis, D. Jovanovski, S. Joordens, R. Mraz, and S. J. Graham, “Utilizing Virtual Reality to Improve the Ecological Validity of Clinical Neuropsychology: An fMRI Case Study Elucidating the Neural Basis of Planning by Comparing the Tower of London with a Three-Dimensional Navigation Task,” *Appl. Neuropsychol.*, vol. 16, no. 4, pp. 295–306, Nov. 2009, doi: 10.1080/09084280903297891.
[58] I. Tarnanas, W. Schlee, M. Tsolaki, R. Müri, U. Mosimann, and T. Nef, “Ecological Validity of Virtual Reality Daily Living Activities Screening for Early Dementia: Longitudinal Study,” *JMIR Serious Games*, vol. 1, no. 1, p. e2778, Aug. 2013, doi: 10.2196/games.2778.
[59] D. M. Cruz-Oliver, T. K. Malmstrom, M. Roegner, N. Tumosa, and G. T. Grossberg, “Cognitive Deficit Reversal as Shown by Changes in the Veterans Affairs Saint Louis University Mental Status (SLUMS) Examination Scores 7.5 Years Later,” *J. Am. Med. Dir. Assoc.*, vol. 15, no. 9, p. 687.e5-687.e10, Sep. 2014, doi: 10.1016/j.jamda.2014.05.004.
[60] “Mini-Mental State Examination - an overview | ScienceDirect Topics.” https://www.sciencedirect.com/topics/medicine-and-dentistry/mini-mental-state-examination (accessed May 10, 2022).
[61] A. J. Mitchell, “The Mini-Mental State Examination (MMSE): Update on Its Diagnostic Accuracy and Clinical Utility for Cognitive Disorders,” in *Cognitive Screening Instruments: A Practical Approach*, A.

J. Larner, Ed. Cham: Springer International Publishing, 2017, pp. 37–48. doi: 10.1007/978-3-319-44775-9_3.
[62] D. P. Seitz *et al.*, "Mini-Cog for the diagnosis of Alzheimer's disease dementia and other dementias within a primary care setting," *Cochrane Database Syst. Rev.*, vol. 2, no. 2, p. CD011415, Feb. 2018, doi: 10.1002/14651858.CD011415.pub2.
[63] mmLearn.org, "Dementia Assessment Tool: The SLUMS Test." https://training.mmlearn.org/blog/dementia-assessment-tool-the-slums-test (accessed Jan. 13, 2023).
[64] D. Szcześniak and J. Rymaszewska, "The usfulness of the SLUMS test for diagnosis of mild cognitive impairment and dementia," *Psychiatr. Pol.*, vol. 50, no. 2, pp. 457–472, 2016, doi: 10.12740/PP/OnlineFirst/43141.
[65] G. Chételat *et al.*, "Amyloid-PET and 18F-FDG-PET in the diagnostic investigation of Alzheimer's disease and other dementias," *Lancet Neurol.*, vol. 19, no. 11, pp. 951–962, Nov. 2020, doi: 10.1016/S1474-4422(20)30314-8.
[66] L. Agüera-Ortiz *et al.*, "Structural correlates of apathy in Alzheimer's disease: a multimodal MRI study," *Int. J. Geriatr. Psychiatry*, vol. 32, no. 8, pp. 922–930, 2017, doi: 10.1002/gps.4548.
[67] J. Peres and A. G. Nasello, "Psychotherapy and neuroscience: Towards closer integration," *Int. J. Psychol.*, vol. 43, no. 6, pp. 943–957, 2008, doi: 10.1080/00207590701248487.
[68] A. Aztiria, M. Ugarte, and A. Izaguirre, "Gait Analysis for Identifying Parameters Related to Dementia in Intelligent Environments," in *2013 9th International Conference on Intelligent Environments*, Jul. 2013, pp. 228–231. doi: 10.1109/IE.2013.10.
[69] X. Sun *et al.*, "Biosensors toward behavior detection in diagnosis of alzheimer's disease," *Front. Bioeng. Biotechnol.*, vol. 10, 2022, Accessed: Jan. 13, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fbioe.2022.1031833
[70] F. Haider, S. de la Fuente, and S. Luz, "An Assessment of Paralinguistic Acoustic Features for Detection of Alzheimer's Dementia in Spontaneous Speech," *IEEE J. Sel. Top. Signal Process.*, vol. 14, no. 2, pp. 272–281, Feb. 2020, doi: 10.1109/JSTSP.2019.2955022.
[71] V. N. Vahia, "Diagnostic and statistical manual of mental disorders 5: A quick glance," *Indian J. Psychiatry*, vol. 55, no. 3, pp. 220–223, 2013, doi: 10.4103/0019-5545.117131.
[72] L.-A. Sapey-Triomphe *et al.*, "Neuroanatomical Correlates of Recognizing Face Expressions in Mild Stages of Alzheimer's Disease," *PloS One*, vol. 10, no. 12, p. e0143586, 2015, doi: 10.1371/journal.pone.0143586.
[73] J. Van den Stock *et al.*, "Impaired recognition of body expressions in the behavioral variant of frontotemporal dementia," *Neuropsychologia*, vol. 75, pp. 496–504, Aug. 2015, doi: 10.1016/j.neuropsychologia.2015.06.035.
[74] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 39–58, Jan. 2009, doi: 10.1109/TPAMI.2008.52.
[75] U. Seidl, U. Lueken, P. A. Thomann, A. Kruse, and J. Schröder, "Facial Expression in Alzheimer's Disease: Impact of Cognitive Deficits and Neuropsychiatric Symptoms," *Am. J. Alzheimers Dis. Dementias®*, vol. 27, no. 2, pp. 100–106, Mar. 2012, doi: 10.1177/1533317512440495.
[76] M. Sethi, S. Ahuja, S. Rani, D. Koundal, A. Zaguia, and W. Enbeyle, "An Exploration: Alzheimer's Disease Classification Based on Convolutional Neural Network," *BioMed Res. Int.*, vol. 2022, p. e8739960, Jan. 2022, doi: 10.1155/2022/8739960.
[77] "Clinical Positron Emission Tomography (PET) Neuroimaging: Advantages and Limitations as a Diagnostic Tool | The Journal of Neuropsychiatry and Clinical Neurosciences." https://neuro.psychiatryonline.org/doi/10.1176/appi.neuropsych.16030044 (accessed Jan. 22, 2023).

[78] *A First Course in Machine Learning*. Chapman and Hall/CRC, 2016. doi: 10.1201/9781315382159.
[79] J. W. M. de Souza, S. S. A. Alves, E. de S. Rebouças, J. S. Almeida, and P. P. Rebouças Filho, "A New Approach to Diagnose Parkinson's Disease Using a Structural Cooccurrence Matrix for a Similarity Analysis," *Comput. Intell. Neurosci.*, vol. 2018, p. e7613282, Apr. 2018, doi: 10.1155/2018/7613282.
[80] E. Pellegrini *et al.*, "Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: A systematic review," *Alzheimers Dement. Diagn. Assess. Dis. Monit.*, vol. 10, pp. 519–535, Aug. 2018, doi: 10.1016/j.dadm.2018.07.004.
[81] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, J. Tohka, and Alzheimer's Disease Neuroimaging Initiative, "Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects," *NeuroImage*, vol. 104, pp. 398–412, Jan. 2015, doi: 10.1016/j.neuroimage.2014.10.002.
[82] H. Bisgin *et al.*, "Comparing SVM and ANN based Machine Learning Methods for Species Identification of Food Contaminating Beetles," *Sci. Rep.*, vol. 8, no. 1, Art. no. 1, Apr. 2018, doi: 10.1038/s41598-018-24926-7.
[83] Z. Shi, L. He, K. Suzuki, T. Nakamura, and H. Itoh, "Survey on Neural Networks Used for Medical Image Processing," *Int. J. Comput. Sci.*, vol. 3, no. 1, pp. 86–100, Feb. 2009.
[84] D. Shen, G. Wu, and H.-I. Suk, "Deep Learning in Medical Image Analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, no. 1, pp. 221–248, 2017, doi: 10.1146/annurev-bioeng-071516-044442.
[85] H.-I. Suk, S.-W. Lee, D. Shen, and Alzheimer's Disease Neuroimaging Initiative, "Deep ensemble learning of sparse regression models for brain disease diagnosis," *Med. Image Anal.*, vol. 37, pp. 101–113, Apr. 2017, doi: 10.1016/j.media.2017.01.008.
[86] M. Oloko-Oba and S. Viriri, "Ensemble of EfficientNets for the Diagnosis of Tuberculosis," *Comput. Intell. Neurosci.*, vol. 2021, p. 9790894, 2021, doi: 10.1155/2021/9790894.
[87] J. Ramírez *et al.*, "Ensemble of random forests One vs. Rest classifiers for MCI and AD prediction using ANOVA cortical and subcortical feature selection and partial least squares," *J. Neurosci. Methods*, vol. 302, pp. 47–57, May 2018, doi: 10.1016/j.jneumeth.2017.12.005.
[88] A. Ortiz, J. Munilla, J. M. Górriz, and J. Ramírez, "Ensembles of Deep Learning Architectures for the Early Diagnosis of the Alzheimer's Disease," *Int. J. Neural Syst.*, Aug. 2016, doi: 10.1142/S0129065716500258.
[89] M. A. Ganaie and M. Tanveer, "Fuzzy least squares projection twin support vector machines for class imbalance learning," *Appl. Soft Comput.*, vol. 113, p. 107933, Dec. 2021, doi: 10.1016/j.asoc.2021.107933.
[90] Y. Tian and Z. Qi, "Review on: Twin Support Vector Machines," *Ann. Data Sci.*, vol. 1, no. 2, pp. 253–277, Jun. 2014, doi: 10.1007/s40745-014-0018-4.
[91] P. Vemuri *et al.*, "Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies," *NeuroImage*, vol. 39, no. 3, pp. 1186–1197, Feb. 2008, doi: 10.1016/j.neuroimage.2007.09.073.
[92] L. Mesrob *et al.*, "Identification of Atrophy Patterns in Alzheimer's Disease Based on SVM Feature Selection and Anatomical Parcellation," in *Medical Imaging and Augmented Reality*, Berlin, Heidelberg, 2008, pp. 124–132. doi: 10.1007/978-3-540-79982-5_14.
[93] B. Magnin *et al.*, "Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI," *Neuroradiology*, vol. 51, no. 2, pp. 73–83, Feb. 2009, doi: 10.1007/s00234-008-0463-x.
[94] E. Gerardin *et al.*, "Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging," *NeuroImage*, vol. 47, no. 4, pp. 1476–1486, Oct. 2009, doi: 10.1016/j.neuroimage.2009.05.036.

[95] F. Zhang, M. Petersen, L. Johnson, J. Hall, and S. E. O'Bryant, "Recursive Support Vector Machine Biomarker Selection for Alzheimer's Disease," *J. Alzheimers Dis. JAD*, vol. 79, no. 4, pp. 1691–1700, 2021, doi: 10.3233/JAD-201254.
[96] N. N. Kulkarni and V. K. Bairagi, "Extracting Salient Features for EEG-based Diagnosis of Alzheimer's Disease Using Support Vector Machine Classifier," *IETE J. Res.*, vol. 63, no. 1, pp. 11–22, Jan. 2017, doi: 10.1080/03772063.2016.1241164.
[97] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.
[98] D. Agarwal, G. Marques, I. de la Torre-Díez, M. A. Franco Martin, B. García Zapiraín, and F. Martín Rodríguez, "Transfer Learning for Alzheimer's Disease through Neuroimaging Biomarkers: A Systematic Review," *Sensors*, vol. 21, no. 21, p. 7259, Oct. 2021, doi: 10.3390/s21217259.
[99] "Grand Challenge," *grand-challenge.org*. https://grand-challenge.org/ (accessed Jan. 16, 2023).
[100] J. Brownlee, "A Gentle Introduction to the ImageNet Challenge (ILSVRC)," *Machine Learning Mastery*, Apr. 30, 2019. https://machinelearningmastery.com/introduction-to-the-imagenet-large-scale-visual-recognition-challenge-ilsvrc/ (accessed May 19, 2022).
[101] D. Agarwal, M. A. Berbis, T. Martín-Noguerol, A. Luna, S. C. P. Garcia, and I. de la Torre-Díez, "End-to-End Deep Learning Architectures Using 3D Neuroimaging Biomarkers for Early Alzheimer's Diagnosis," *Mathematics*, vol. 10, no. 15, Art. no. 15, Jan. 2022, doi: 10.3390/math10152575.
[102] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015, doi: 10.1016/j.neunet.2014.09.003.
[103] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A Fast Learning Algorithm for Deep Belief Nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006, doi: 10.1162/neco.2006.18.7.1527.
[104] T. Uemura, J. J. Näppi, T. Hironaka, H. Kim, and H. Yoshida, "Comparative performance of 3D-DenseNet, 3D-ResNet, and 3D-VGG models in polyp detection for CT colonography," in *Medical Imaging 2020: Computer-Aided Diagnosis*, Mar. 2020, vol. 11314, pp. 736–741. doi: 10.1117/12.2549103.
[105] "A Fast Learning Algorithm for Deep Belief Nets | Neural Computation | MIT Press." https://direct.mit.edu/neco/article-abstract/18/7/1527/7065/A-Fast-Learning-Algorithm-for-Deep-Belief-Nets?redirectedFrom=fulltext (accessed Jan. 16, 2023).
[106] M. Tanveer *et al.*, "Machine Learning Techniques for the Diagnosis of Alzheimer's Disease: A Review," *ACM Trans. Multimed. Comput. Commun. Appl.*, vol. 16, no. 1s, p. 30:1-30:35, Apr. 2020, doi: 10.1145/3344998.
[107] R. Li *et al.*, "Deep learning based imaging data completion for improved brain disease diagnosis," *Med. Image Comput. Comput.-Assist. Interv. MICCAI Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, vol. 17, no. Pt 3, pp. 305–312, 2014, doi: 10.1007/978-3-319-10443-0_39.
[108] D. Lu, K. Popuri, G. W. Ding, R. Balachandar, M. F. Beg, and Alzheimer's Disease Neuroimaging Initiative, "Multimodal and Multiscale Deep Neural Networks for the Early Diagnosis of Alzheimer's Disease using structural MR and FDG-PET images," *Sci. Rep.*, vol. 8, no. 1, p. 5697, Apr. 2018, doi: 10.1038/s41598-018-22871-z.
[109] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Landmark-based deep multi-instance learning for brain disease diagnosis," *Med. Image Anal.*, vol. 43, pp. 157–168, Jan. 2018, doi: 10.1016/j.media.2017.10.005.
[110] H.-I. Suk, S.-W. Lee, D. Shen, and The Alzheimer's Disease Neuroimaging Initiative, "Latent feature representation with stacked auto-encoder for AD/MCI diagnosis," *Brain Struct. Funct.*, vol. 220, no. 2, pp. 841–859, Mar. 2015, doi: 10.1007/s00429-013-0687-3.
[111] H.-I. Suk, S.-W. Lee, and D. Shen, "Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis," *NeuroImage*, vol. 101, pp. 569–582, Nov. 2014, doi: 10.1016/j.neuroimage.2014.06.077.

[112] F. Li, L. Tran, K.-H. Thung, S. Ji, D. Shen, and J. Li, “A Robust Deep Model for Improved Classification of AD/MCI Patients,” *IEEE J. Biomed. Health Inform.*, vol. 19, no. 5, pp. 1610–1616, Sep. 2015, doi: 10.1109/JBHI.2015.2429556.
[113] M. J. Page *et al.*, “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews,” *BMJ*, p. n71, Mar. 2021, doi: 10.1136/bmj.n71.
[114] M. B. Miles and A. M. Huberman, *Qualitative Data Analysis: An Expanded Sourcebook*. SAGE, 1994.
[115] “SMCI vs AD | Sysrev.” https://sysrev.com/u/2642/p/118998/article/14772257 (accessed Apr. 18, 2022).
[116] A. Shakarami, H. Tarrah, and A. Mahdavi-Hormat, “A CAD system for diagnosing Alzheimer’s disease using 2D slices and an improved AlexNet-SVM method,” *Optik*, vol. 212, p. 164237, Jun. 2020, doi: 10.1016/j.ijleo.2020.164237.
[117] C. Ieracitano, N. Mammone, A. Bramanti, A. Hussain, and F. C. Morabito, “A Convolutional Neural Network approach for classification of dementia stages based on 2D-spectral representation of EEG recordings,” *Neurocomputing*, vol. 323, pp. 96–107, Jan. 2019, doi: 10.1016/j.neucom.2018.09.071.
[118] S. Basheera and M. Satya Sai Ram, “A novel CNN based Alzheimer’s disease classification using hybrid enhanced ICA segmented gray matter of MRI,” *Comput. Med. Imaging Graph.*, vol. 81, p. 101713, Apr. 2020, doi: 10.1016/j.compmedimag.2020.101713.
[119] H. Shahamat and M. Saniee Abadeh, “Brain MRI analysis using a deep learning based evolutionary approach,” *Neural Netw.*, vol. 126, pp. 218–234, Jun. 2020, doi: 10.1016/j.neunet.2020.03.017.
[120] C.-Y. Wee *et al.*, “Cortical graph neural network for AD and MCI diagnosis and transfer learning across populations,” *NeuroImage Clin.*, vol. 23, p. 101929, 2019, doi: 10.1016/j.nicl.2019.101929.
[121] A. B. Tufail, Y. Ma, and Q.-N. Zhang, “Multiclass classification of initial stages of Alzheimer’s Disease through Neuroimaging modalities and Convolutional Neural Networks,” in *2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*, Jun. 2020, pp. 51–56. doi: 10.1109/ITOEC49072.2020.9141553.
[122] X. Bi, S. Li, B. Xiao, Y. Li, G. Wang, and X. Ma, “Computer aided Alzheimer’s disease diagnosis by an unsupervised deep learning technology,” *Neurocomputing*, vol. 392, pp. 296–304, Jun. 2020, doi: 10.1016/j.neucom.2018.11.111.
[123] “Convolution neural network–based Alzheimer’s disease classification using hybrid enhanced independent component analysis based segmented gray matter of T2 weighted magnetic resonance imaging with clinical valuation - Basheera - 2019 - Alzheimer’s & Dementia: Translational Research & Clinical Interventions - Wiley Online Library.” https://alz-journals.onlinelibrary.wiley.com/doi/full/10.1016/j.trci.2019.10.001 (accessed Jan. 19, 2023).
[124] W. Li, X. Lin, and X. Chen, “Detecting Alzheimer’s disease Based on 4D fMRI: An exploration under deep learning framework,” *Neurocomputing*, vol. 388, pp. 280–287, May 2020, doi: 10.1016/j.neucom.2020.01.053.
[125] H. Wang *et al.*, “Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer’s disease,” *Neurocomputing*, vol. 333, pp. 145–156, Mar. 2019, doi: 10.1016/j.neucom.2018.12.018.
[126] S. Qiu, G. H. Chang, M. Panagia, D. M. Gopal, R. Au, and V. B. Kolachalama, “Fusion of deep learning models of MRI scans, Mini–Mental State Examination, and logical memory test enhances diagnosis of mild cognitive impairment,” *Alzheimers Dement. Diagn. Assess. Dis. Monit.*, vol. 10, no. 1, pp. 737–749, 2018, doi: 10.1016/j.dadm.2018.08.013.

[127] F. Zhang, Z. Li, B. Zhang, H. Du, B. Wang, and X. Zhang, “Multi-modal deep learning model for auxiliary diagnosis of Alzheimer’s disease,” *Neurocomputing*, vol. 361, pp. 185–195, Oct. 2019, doi: 10.1016/j.neucom.2019.04.093.
[128] K. R. Kruthika, Rajeswari, and H. D. Maheshappa, “CBIR system using Capsule Networks and 3D CNN for Alzheimer’s disease diagnosis,” *Inform. Med. Unlocked*, vol. 14, pp. 59–68, Jan. 2019, doi: 10.1016/j.imu.2018.12.001.
[129] M. Rohini and D. Surendran, “Classification of Neurodegenerative Disease Stages using Ensemble Machine Learning Classifiers,” *Procedia Comput. Sci.*, vol. 165, pp. 66–73, Jan. 2019, doi: 10.1016/j.procs.2020.01.071.
[130] B. Zhang, J. Wang, L. Lin, and S. Wu, “The Effect of Smoothing Filter on CNN based AD Classification,” in *Proceedings of the 2019 8th International Conference on Bioinformatics and Biomedical Science*, New York, NY, USA, Jan. 2020, pp. 54–58. doi: 10.1145/3369166.3369191.
[131] N. M. Khan, N. Abraham, and M. Hon, “Transfer Learning With Intelligent Training Data Selection for Prediction of Alzheimer’s Disease,” *IEEE Access*, vol. 7, pp. 72726–72735, 2019, doi: 10.1109/ACCESS.2019.2920448.
[132] “Transfer Learning With Intelligent Training Data Selection for Prediction of Alzheimer’s Disease | IEEE Journals & Magazine | IEEE Xplore.” https://ieeexplore.ieee.org/document/8727911 (accessed Jan. 19, 2023).
[133] M. Liu *et al.*, “A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer’s disease,” *NeuroImage*, vol. 208, p. 116459, Mar. 2020, doi: 10.1016/j.neuroimage.2019.116459.
[134] H. Ji, Z. Liu, W. Q. Yan, and R. Klette, “Early Diagnosis of Alzheimer’s Disease Using Deep Learning,” in *Proceedings of the 2nd International Conference on Control and Computer Vision*, New York, NY, USA, Jun. 2019, pp. 87–91. doi: 10.1145/3341016.3341024.
[135] F. Gao, H. Yoon, T. Wu, and X. Chu, “A feature transfer enabled multi-task deep learning model on medical imaging,” *Expert Syst. Appl.*, vol. 143, p. 112957, Apr. 2020, doi: 10.1016/j.eswa.2019.112957.
[136] E. Yagis, A. G. S. De Herrera, and L. Citi, “Generalization Performance of Deep Learning Models in Neurodegenerative Disease Classification,” in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Nov. 2019, pp. 1692–1698. doi: 10.1109/BIBM47256.2019.8983088.
[137] F. Ramzan *et al.*, “A Deep Learning Approach for Automated Diagnosis and Multi-Class Classification of Alzheimer’s Disease Stages Using Resting-State fMRI and Residual Neural Networks,” *J. Med. Syst.*, vol. 44, no. 2, p. 37, Dec. 2019, doi: 10.1007/s10916-019-1475-2.
[138] F. Li and M. Liu, “A hybrid Convolutional and Recurrent Neural Network for Hippocampus Analysis in Alzheimer’s Disease,” *J. Neurosci. Methods*, vol. 323, pp. 108–118, Jul. 2019, doi: 10.1016/j.jneumeth.2019.05.006.
[139] F. Gao *et al.*, “AD-NET: Age-adjust neural network for improved MCI to AD conversion prediction,” *NeuroImage Clin.*, vol. 27, p. 102290, Jan. 2020, doi: 10.1016/j.nicl.2020.102290.
[140] S. Basaia *et al.*, “Automated classification of Alzheimer’s disease and mild cognitive impairment using a single MRI and deep neural networks,” *NeuroImage Clin.*, vol. 21, p. 101645, 2019, doi: 10.1016/j.nicl.2018.101645.
[141] K. Oh, Y.-C. Chung, K. W. Kim, W.-S. Kim, and I.-S. Oh, “Classification and Visualization of Alzheimer’s Disease using Volumetric Convolutional Neural Network and Transfer Learning,” *Sci. Rep.*, vol. 9, no. 1, Art. no. 1, Dec. 2019, doi: 10.1038/s41598-019-54548-6.
[142] J. Wen *et al.*, “Convolutional neural networks for classification of Alzheimer’s disease: Overview and reproducible evaluation,” *Med. Image Anal.*, vol. 63, p. 101694, Jul. 2020, doi: 10.1016/j.media.2020.101694.

[143] L. Lin and B. Zhang, “MCI Conversion Prediction Based on Transfer Learning,” *DEStech Trans. Comput. Sci. Eng.*, vol. 0, no. CCNT, Art. no. CCNT, 2018, doi: 10.12783/dtcse/CCNT2018/24702.
[144] D. Lu, K. Popuri, G. W. Ding, R. Balachandar, M. F. Beg, and Alzheimer’s Disease Neuroimaging Initiative, “Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer’s disease,” *Med. Image Anal.*, vol. 46, pp. 26–34, May 2018, doi: 10.1016/j.media.2018.02.002.
[145] H. Choi, K. H. Jin, and Alzheimer’s Disease Neuroimaging Initiative, “Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging,” *Behav. Brain Res.*, vol. 344, pp. 103–109, May 2018, doi: 10.1016/j.bbr.2018.02.017.
[146] Z. Yang and Z. Liu, “The risk prediction of Alzheimer’s disease based on the deep learning model of brain 18F-FDG positron emission tomography,” *Saudi J. Biol. Sci.*, vol. 27, no. 2, pp. 659–665, Feb. 2020, doi: 10.1016/j.sjbs.2019.12.004.
[147] M. Hon and N. M. Khan, “Towards Alzheimer’s disease classification through transfer learning,” in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Nov. 2017, pp. 1166–1169. doi: 10.1109/BIBM.2017.8217822.
[148] G. E. Hinton and R. R. Salakhutdinov, “Reducing the Dimensionality of Data with Neural Networks,” *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006, doi: 10.1126/science.1127647.
[149] R. Salakhutdinov and H. Larochelle, “Efficient Learning of Deep Boltzmann Machines,” in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Mar. 2010, pp. 693–700. Accessed: Jan. 21, 2023. [Online]. Available: https://proceedings.mlr.press/v9/salakhutdinov10a.html
[150] W. Samek, A. Binder, G. Montavon, S. Lapuschkin, and K.-R. Müller, “Evaluating the Visualization of What a Deep Neural Network Has Learned,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 11, pp. 2660–2673, Nov. 2017, doi: 10.1109/TNNLS.2016.2599820.
[151] F. Gaillard, “MRI sequences (overview) | Radiology Reference Article | Radiopaedia.org,” *Radiopaedia*. https://radiopaedia.org/articles/mri-sequences-overview (accessed Dec. 10, 2022).
[152] P. Vemuri and C. R. Jack, “Role of structural MRI in Alzheimer’s disease,” *Alzheimers Res. Ther.*, vol. 2, no. 4, p. 23, Aug. 2010, doi: 10.1186/alzrt47.
[153] E. L. Dennis and P. M. Thompson, “Functional Brain Connectivity Using fMRI in Aging and Alzheimer’s Disease,” *Neuropsychol. Rev.*, vol. 24, no. 1, pp. 49–62, Mar. 2014, doi: 10.1007/s11065-014-9249-6.
[154] W. Pelkmans *et al.*, “Gray matter T1-w/T2-w ratios are higher in Alzheimer’s disease,” *Hum. Brain Mapp.*, vol. 40, no. 13, pp. 3900–3909, Jun. 2019, doi: 10.1002/hbm.24638.
[155] M. Brant-Zawadzki, G. D. Gillan, and W. R. Nitz, “MP RAGE: a three-dimensional, T1-weighted, gradient-echo sequence--initial experience in the brain.,” *Radiology*, vol. 182, no. 3, pp. 769–775, Mar. 1992, doi: 10.1148/radiology.182.3.1535892.
[156] T. Jo, K. Nho, S. L. Risacher, A. J. Saykin, and for the Alzheimer’s Neuroimaging Initiative, “Deep learning detection of informative features in tau PET for Alzheimer’s disease classification,” *BMC Bioinformatics*, vol. 21, no. 21, p. 496, Dec. 2020, doi: 10.1186/s12859-020-03848-0.
[157] Y.-N. Ou *et al.*, “FDG-PET as an independent biomarker for Alzheimer’s biological diagnosis: a longitudinal study,” *Alzheimers Res. Ther.*, vol. 11, no. 1, p. 57, Jun. 2019, doi: 10.1186/s13195-019-0512-1.
[158] L. Saint-Aubert, L. Lemoine, K. Chiotis, A. Leuzy, E. Rodriguez-Vieitez, and A. Nordberg, “Tau PET imaging: present and future directions,” *Mol. Neurodegener.*, vol. 12, no. 1, p. 19, Feb. 2017, doi: 10.1186/s13024-017-0162-3.
[159] E. L. Werry *et al.*, “Recent Developments in TSPO PET Imaging as A Biomarker of Neuroinflammation in Neurodegenerative Disorders,” *Int. J. Mol. Sci.*, vol. 20, no. 13, p. 3161, Jun. 2019, doi: 10.3390/ijms20133161.

[160] Z. Cai, S. Li, D. Matuskey, N. Nabulsi, and Y. Huang, “PET imaging of synaptic density: A new tool for investigation of neuropsychiatric diseases,” *Neurosci. Lett.*, vol. 691, pp. 44–50, Jan. 2019, doi: 10.1016/j.neulet.2018.07.038.
[161] X. Sun *et al.*, “Histogram-based normalization technique on human brain magnetic resonance images from different acquisitions,” *Biomed. Eng. Online*, vol. 14, p. 73, Jul. 2015, doi: 10.1186/s12938-015-0064-y.
[162] J. Ashburner, “A fast diffeomorphic image registration algorithm,” *NeuroImage*, vol. 38, no. 1, pp. 95–113, Oct. 2007, doi: 10.1016/j.neuroimage.2007.07.007.
[163] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, “Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain,” *Med. Image Anal.*, vol. 12, no. 1, pp. 26–41, Feb. 2008, doi: 10.1016/j.media.2007.06.004.
[164] J. Ashburner and K. J. Friston, “Unified segmentation,” *NeuroImage*, vol. 26, no. 3, pp. 839–851, Jul. 2005, doi: 10.1016/j.neuroimage.2005.02.018.
[165] S.-H. Choi, J.-G. Chi, Y.-B. Kim, and Z.-H. Cho, “Anterior Commissure - Posterior Commissure Revisited,” *Korean J. Radiol.*, vol. 14, no. 4, pp. 653–661, 2013, doi: 10.3348/kjr.2013.14.4.653.
[166] D. C. Newitt *et al.*, “Gradient Nonlinearity Correction to Improve Apparent Diffusion Coefficient Accuracy and Standardization in the American College of Radiology Imaging Network 6698 Breast Cancer Trial,” *J. Magn. Reson. Imaging JMRI*, vol. 42, no. 4, pp. 908–919, Oct. 2015, doi: 10.1002/jmri.24883.
[167] J. Juntu, J. Sijbers, D. Van Dyck, and J. Gielen, “Bias Field Correction for MRI Images,” in *Computer Recognition Systems*, Berlin, Heidelberg, 2005, pp. 543–551. doi: 10.1007/3-540-32390-2_64.
[168] N. J. Tustison *et al.*, “N4ITK: improved N3 bias correction,” *IEEE Trans. Med. Imaging*, vol. 29, no. 6, pp. 1310–1320, Jun. 2010, doi: 10.1109/TMI.2010.2046908.
[169] A. Li, F. Li, F. Elahifasaee, M. Liu, L. Zhang, and Alzheimer’s Disease Neuroimaging Initiative, “Hippocampal shape and asymmetry analysis by cascaded convolutional neural networks for Alzheimer’s disease diagnosis,” *Brain Imaging Behav.*, vol. 15, no. 5, pp. 2330–2339, Oct. 2021, doi: 10.1007/s11682-020-00427-y.
[170] R. Cui and M. Liu, “Hippocampus Analysis by Combination of 3-D DenseNet and Shapes for Alzheimer’s Disease Diagnosis,” *IEEE J. Biomed. Health Inform.*, vol. 23, no. 5, pp. 2099–2107, Sep. 2019, doi: 10.1109/JBHI.2018.2882392.
[171] X. Long, L. Chen, C. Jiang, L. Zhang, and Alzheimer’s Disease Neuroimaging Initiative, “Prediction and classification of Alzheimer disease based on quantification of MRI deformation,” *PloS One*, vol. 12, no. 3, p. e0173372, 2017, doi: 10.1371/journal.pone.0173372.
[172] N. J. Kabani, D. J. MacDonald, C. J. Holmes, and A. C. Evans, “3D Anatomical Atlas of the Human Brain,” *NeuroImage*, vol. 7, no. 4, Part 2, p. S717, May 1998, doi: 10.1016/S1053-8119(18)31550-7.
[173] E. T. Rolls, C.-C. Huang, C.-P. Lin, J. Feng, and M. Joliot, “Automated anatomical labelling atlas 3,” *NeuroImage*, vol. 206, p. 116189, Feb. 2020, doi: 10.1016/j.neuroimage.2019.116189.
[174] M. D. Zeiler and R. Fergus, “Visualizing and Understanding Convolutional Networks,” in *Computer Vision – ECCV 2014*, Cham, 2014, pp. 818–833. doi: 10.1007/978-3-319-10590-1_53.
[175] C. Shorten and T. M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning,” *J. Big Data*, vol. 6, no. 1, p. 60, Jul. 2019, doi: 10.1186/s40537-019-0197-0.
[176] A. Mikołajczyk and M. Grochowski, “Data augmentation for improving deep learning in image classification problem,” in *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, May 2018, pp. 117–122. doi: 10.1109/IIPHDW.2018.8388338.
[177] L. Perez and J. Wang, “The Effectiveness of Data Augmentation in Image Classification using Deep Learning.” arXiv, Dec. 13, 2017. doi: 10.48550/arXiv.1712.04621.

[178] F. J. Martinez-Murcia *et al.*, “Functional Brain Imaging Synthesis Based on Image Decomposition and Kernel Modeling: Application to Neurodegenerative Diseases,” *Front. Neuroinformatics*, vol. 11, p. 65, 2017, doi: 10.3389/fninf.2017.00065.
[179] M. Mardani *et al.*, “Deep Generative Adversarial Networks for Compressed Sensing Automates MRI.” arXiv, May 31, 2017. doi: 10.48550/arXiv.1706.00051.
[180] H.-C. Shin *et al.*, “Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning,” *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1285–1298, May 2016, doi: 10.1109/TMI.2016.2528162.
[181] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” *ArXiv160806993 Cs*, Jan. 2018, Accessed: May 05, 2022. [Online]. Available: http://arxiv.org/abs/1608.06993
[182] T. Glasmachers, “Limits of End-to-End Learning,” in *Proceedings of the Ninth Asian Conference on Machine Learning*, Nov. 2017, pp. 17–32. Accessed: Apr. 30, 2022. [Online]. Available: https://proceedings.mlr.press/v77/glasmachers17a.html
[183] S. Vieira, W. H. L. Pinaya, and A. Mechelli, “Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications,” *Neurosci. Biobehav. Rev.*, vol. 74, no. Pt A, pp. 58–75, Mar. 2017, doi: 10.1016/j.neubiorev.2017.01.002.
[184] M. Liu, J. Zhang, C. Lian, and D. Shen, “Weakly Supervised Deep Learning for Brain Disease Prognosis Using MRI and Incomplete Clinical Scores,” *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3381–3392, Jul. 2020, doi: 10.1109/TCYB.2019.2904186.
[185] S. Toshkhujaev *et al.*, “Classification of Alzheimer’s Disease and Mild Cognitive Impairment Based on Cortical and Subcortical Features from MRI T1 Brain Images Utilizing Four Different Types of Datasets,” *J. Healthc. Eng.*, vol. 2020, p. e3743171, Sep. 2020, doi: 10.1155/2020/3743171.
[186] R. Ju, C. Hu, pan zhou, and Q. Li, “Early Diagnosis of Alzheimer’s Disease Based on Resting-State Brain Networks and Deep Learning,” *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 16, no. 1, pp. 244–257, Jan. 2019, doi: 10.1109/TCBB.2017.2776910.
[187] M. Nguyen *et al.*, “Predicting Alzheimer’s disease progression using deep recurrent neural networks,” *NeuroImage*, vol. 222, p. 117203, Nov. 2020, doi: 10.1016/j.neuroimage.2020.117203.
[188] G. Folego, M. Weiler, R. F. Casseb, R. Pires, and A. Rocha, “Alzheimer’s Disease Detection Through Whole-Brain 3D-CNN MRI,” *Front. Bioeng. Biotechnol.*, vol. 8, 2020, Accessed: Apr. 25, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/fbioe.2020.534592
[189] S. Basheera and M. S. Sai Ram, “Convolution neural network-based Alzheimer’s disease classification using hybrid enhanced independent component analysis based segmented gray matter of T2 weighted magnetic resonance imaging with clinical valuation,” *Alzheimers Dement. N. Y. N*, vol. 5, pp. 974–986, 2019, doi: 10.1016/j.trci.2019.10.001.
[190] A. Mehmood *et al.*, “A Transfer Learning Approach for Early Diagnosis of Alzheimer’s Disease on MRI Images,” *Neuroscience*, vol. 460, pp. 43–52, Apr. 2021, doi: 10.1016/j.neuroscience.2021.01.002.
[191] T. A. Tuan, T. B. Pham, J. Y. Kim, and J. M. R. S. Tavares, “Alzheimer’s diagnosis using deep learning in segmenting and classifying 3D brain MR images,” *Int. J. Neurosci.*, pp. 1–10, Nov. 2020, doi: 10.1080/00207454.2020.1835900.
[192] J. Song, J. Zheng, P. Li, X. Lu, G. Zhu, and P. Shen, “An Effective Multimodal Image Fusion Method Using MRI and PET for Alzheimer’s Disease Diagnosis,” *Front. Digit. Health*, vol. 3, 2021, Accessed: Apr. 22, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/fdgth.2021.637386
[193] M. Odusami, R. Maskeliūnas, and R. Damaševičius, “An Intelligent System for Early Recognition of Alzheimer’s Disease Using Neuroimaging,” *Sensors*, vol. 22, no. 3, p. 740, Jan. 2022, doi: 10.3390/s22030740.

[194] X. Bi, W. Liu, H. Liu, and Q. Shang, “Artificial Intelligence-based MRI Images for Brain in Prediction of Alzheimer’s Disease,” *J. Healthc. Eng.*, vol. 2021, p. 8198552, 2021, doi: 10.1155/2021/8198552.
[195] A. Puente-Castro, E. Fernandez-Blanco, A. Pazos, and C. R. Munteanu, “Automatic assessment of Alzheimer’s disease diagnosis based on deep learning techniques,” *Comput. Biol. Med.*, vol. 120, p. 103764, May 2020, doi: 10.1016/j.compbiomed.2020.103764.
[196] A. B. Tufail, Y.-K. Ma, and Q.-N. Zhang, “Binary Classification of Alzheimer’s Disease Using sMRI Imaging Modality and Deep Learning,” *J. Digit. Imaging*, vol. 33, no. 5, pp. 1073–1090, Oct. 2020, doi: 10.1007/s10278-019-00265-5.
[197] N. J. Herzog and G. D. Magoulas, “Brain Asymmetry Detection and Machine Learning Classification for Diagnosis of Early Dementia,” *Sensors*, vol. 21, no. 3, p. 778, Jan. 2021, doi: 10.3390/s21030778.
[198] L. Nanni *et al.*, “Comparison of Transfer Learning and Conventional Machine Learning Applied to Structural Brain MRI for the Early Diagnosis and Prognosis of Alzheimer’s Disease,” *Front. Neurol.*, vol. 11, 2020, Accessed: Apr. 25, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/fneur.2020.576194
[199] J. Jiang, L. Kang, J. Huang, and T. Zhang, “Deep learning based mild cognitive impairment diagnosis using structure MR images,” *Neurosci. Lett.*, vol. 730, p. 134791, Jun. 2020, doi: 10.1016/j.neulet.2020.134971.
[200] D. Prakash, N. Madusanka, S. Bhattacharjee, C.-H. Kim, H.-G. Park, and H.-K. Choi, “Diagnosing Alzheimer’s Disease Based on Multiclass MRI Scans using Transfer Learning Techniques,” *Curr. Med. Imaging*, vol. 17, no. 12, pp. 1460–1472, 2021, doi: 10.2174/1573405617666210127161812.
[201] Y. Gupta *et al.*, “Early diagnosis of Alzheimer’s disease using combined features from voxel-based morphometry and cortical, subcortical, and hippocampus regions of MRI T1 brain images,” *PLOS ONE*, vol. 14, no. 10, p. e0222446, Oct. 2019, doi: 10.1371/journal.pone.0222446.
[202] A. Zeng, L. Jia, D. Pan, and X. Song, “[Early prognosis of Alzheimer’s disease based on convolutional neural networks and ensemble learning],” *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi J. Biomed. Eng. Shengwu Yixue Gongchengxue Zazhi*, vol. 36, no. 5, pp. 711–719, Oct. 2019, doi: 10.7507/1001-5515.201809040.
[203] T. Zhang and M. Shi, “Multi-modal neuroimaging feature fusion for diagnosis of Alzheimer’s disease,” *J. Neurosci. Methods*, vol. 341, p. 108795, Jul. 2020, doi: 10.1016/j.jneumeth.2020.108795.
[204] M. Liu, D. Cheng, K. Wang, Y. Wang, and Alzheimer’s Disease Neuroimaging Initiative, “Multi-Modality Cascaded Convolutional Neural Networks for Alzheimer’s Disease Diagnosis,” *Neuroinformatics*, vol. 16, no. 3–4, pp. 295–308, Oct. 2018, doi: 10.1007/s12021-018-9370-4.
[205] L. Xu, X. Wu, K. Chen, and L. Yao, “Multi-modality sparse representation-based classification for Alzheimer’s disease and mild cognitive impairment,” *Comput. Methods Programs Biomed.*, vol. 122, no. 2, pp. 182–190, Nov. 2015, doi: 10.1016/j.cmpb.2015.08.004.
[206] X. Pan *et al.*, “Multi-view Separable Pyramid Network for AD Prediction at MCI Stage by 18F-FDG Brain PET Imaging,” *IEEE Trans. Med. Imaging*, pp. 1–1, 2020, doi: 10.1109/TMI.2020.3022591.
[207] A. Abrol, Z. Fu, Y. Du, and V. D. Calhoun, “Multimodal Data Fusion of Deep Learning and Dynamic Functional Connectivity Features to Predict Alzheimer’s Disease Progression,” *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.*, vol. 2019, pp. 4409–4413, Jul. 2019, doi: 10.1109/EMBC.2019.8856500.
[208] J. Venugopalan, L. Tong, H. R. Hassanzadeh, and M. D. Wang, “Multimodal deep learning models for early detection of Alzheimer’s disease stage,” *Sci. Rep.*, vol. 11, no. 1, p. 3254, Feb. 2021, doi: 10.1038/s41598-020-74399-w.

[209] J. Shi, X. Zheng, Y. Li, Q. Zhang, and S. Ying, "Multimodal Neuroimaging Feature Learning With Multimodal Stacked Deep Polynomial Networks for Diagnosis of Alzheimer's Disease," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 1, pp. 173–183, Jan. 2018, doi: 10.1109/JBHI.2017.2655720.

[210] T. Shen *et al.*, "Predicting Alzheimer Disease From Mild Cognitive Impairment With a Deep Belief Network Based on 18F-FDG-PET Images," *Mol. Imaging*, vol. 18, p. 1536012119877285, Dec. 2019, doi: 10.1177/1536012119877285.

[211] G. Lee, K. Nho, B. Kang, K.-A. Sohn, and D. Kim, "Predicting Alzheimer's disease progression using multi-modal deep learning approach," *Sci. Rep.*, vol. 9, no. 1, Art. no. 1, Feb. 2019, doi: 10.1038/s41598-018-37769-z.

[212] F. Er and D. Goularas, "Predicting the Prognosis of MCI Patients Using Longitudinal MRI Data," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 18, no. 3, pp. 1164–1173, Jun. 2021, doi: 10.1109/TCBB.2020.3017872.

[213] J. Rieke, F. Eitel, M. Weygandt, J.-D. Haynes, and K. Ritter, "Visualizing Convolutional Networks for MRI-based Diagnosis of Alzheimer's Disease," *ArXiv180802874 Cs*, vol. 11038, pp. 24–31, 2018, doi: 10.1007/978-3-030-02628-8_3.

[214] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3D brain MRI classification," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, Apr. 2017, pp. 835–838. doi: 10.1109/ISBI.2017.7950647.

[215] N. J. Tustison *et al.*, "Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements," *NeuroImage*, vol. 99, pp. 166–179, Oct. 2014, doi: 10.1016/j.neuroimage.2014.05.044.

[216] B. Park, K. Byeon, and H. Park, "FuNP (Fusion of Neuroimaging Preprocessing) Pipelines: A Fully Automated Preprocessing Software for Functional Magnetic Resonance Imaging," *Front. Neuroinformatics*, vol. 13, 2019, Accessed: May 14, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/fninf.2019.00005

[217] N. Bhagwat *et al.*, "Understanding the impact of preprocessing pipelines on neuroimaging cortical surface analyses," *GigaScience*, vol. 10, no. 1, p. giaa155, Jan. 2021, doi: 10.1093/gigascience/giaa155.

[218] J. V. Manjón, P. Coupé, L. Martí-Bonmatí, D. L. Collins, and M. Robles, "Adaptive non-local means denoising of MR images with spatially varying noise levels," *J. Magn. Reson. Imaging JMRI*, vol. 31, no. 1, pp. 192–203, Jan. 2010, doi: 10.1002/jmri.22003.

[219] K. Köser, "Affine Registration," in *Computer Vision: A Reference Guide*, K. Ikeuchi, Ed. Boston, MA: Springer US, 2014, pp. 22–25. doi: 10.1007/978-0-387-31439-6_122.

[220] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *ArXiv190511946 Cs Stat*, Sep. 2020, Accessed: May 05, 2022. [Online]. Available: http://arxiv.org/abs/1905.11946

[221] T. D. Pham, "A comprehensive study on classification of COVID-19 on computed tomography with pretrained convolutional neural networks," *Sci. Rep.*, vol. 10, p. 16942, Oct. 2020, doi: 10.1038/s41598-020-74164-z.

[222] P. Capainolo and R. M. Chase, "An Early Clinical Case of COVID-19 in New York," *Case Rep. Clin. Med.*, vol. 11, no. 08, pp. 330–336, 2022, doi: 10.4236/crcm.2022.118046.

[223] K. Ali, Z. A. Shaikh, A. A. Khan, and A. A. Laghari, "Multiclass skin cancer classification using EfficientNets – a first step towards preventing skin cancer," *Neurosci. Inform.*, vol. 2, no. 4, p. 100034, Dec. 2022, doi: 10.1016/j.neuri.2021.100034.

[224] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," arXiv, arXiv:1412.6980, Jan. 2017. doi: 10.48550/arXiv.1412.6980.

[225] "Adam Optimizer PyTorch With Examples - Python Guides," Feb. 26, 2022. https://pythonguides.com/adam-optimizer-pytorch/ (accessed May 23, 2022).

[226] K. Zorlu, C. Gokceoglu, F. Ocakoglu, H. A. Nefeslioglu, and S. Acikalin, “Prediction of uniaxial compressive strength of sandstones using petrography-based models,” *Eng. Geol.*, vol. 96, no. 3, pp. 141–158, Feb. 2008, doi: 10.1016/j.enggeo.2007.10.009.
[227] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, “Ensemble deep learning: A review,” *Eng. Appl. Artif. Intell.*, vol. 115, p. 105151, Oct. 2022, doi: 10.1016/j.engappai.2022.105151.
[228] D. R. Hess, “Retrospective studies and chart reviews,” *Respir. Care*, vol. 49, no. 10, pp. 1171–1174, Oct. 2004.