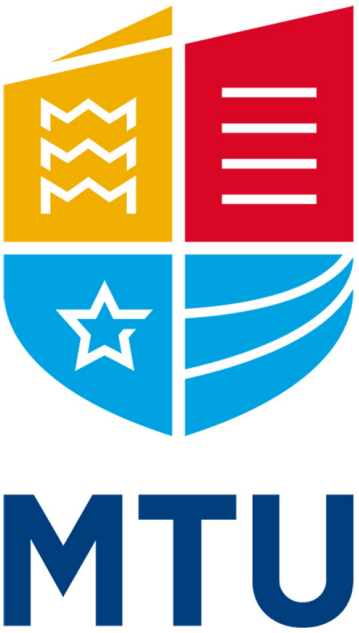

Predicting Alzheimer's using Multimodal Deep Learning

by

Alannah Cullinane Cooney

This thesis has been submitted in partial fulfillment for the degree of Bachelor of Science in Software Development

in the
Faculty of Engineering and Science
Department of Computer Science

April 2024

Declaration of Authorship

This report, Predicting Alzheimer's using Multimodal Deep Learning, is submitted in partial fulfillment of the requirements of Bachelor of Science in Software Development at Munster Technological University Cork. I, Alannah Cullinane Cooney, declare that this thesis titled, Predicting Alzheimer's using Multimodal Deep Learning and the work represents substantially the result of my own work except where explicitly indicated in the text. This report may be freely copied and distributed provided the source is explicitly acknowledged. I confirm that:

- This work was done wholly or mainly while in candidature Bachelor of Science in Software Development at Munster Technological University Cork.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at Munster Technological University Cork or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this project report is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: Alannah Cullinane Cooney

Date: 26/04/2024

MUNSTER TECHNOLOGICAL UNIVERSITY CORK

Abstract

Faculty of Engineering and Science
Department of Computer Science

Bachelor of Science

by Alannah Cullinane Cooney

The goal of this research project is to create a deep learning multi-modal system that will improve the detection and treatment of Alzheimer's disease. The main goal is to forecast Alzheimer's disease in its early stages by using neuroimaging and cerebrospinal fluid (CSF) datasets, which will bring about a paradigm shift in patient management. The system's ultimate goal is to improve patient outcomes by achieving extraordinary performance, accuracy, scalability, and security. The research, which has a strong emphasis on early identification, tackles the urgent need for prompt therapies, enhancing the quality of life for those who are affected and maybe reducing the social burden of Alzheimer's.

The main goals are to improve the deep learning model's performance to accomplish early detection, which is an essential first step in diagnosing, treating, and eventually curing Alzheimer's disease. Aiming for a fine balance between specificity and sensitivity, the research prioritizes accuracy to spare patients and carers needless anxiety. The research introduces a multi-modal accuracy technique that captures a comprehensive view of brain structure and function by merging data from neuroimaging and CSF datasets. The early detection and treatment of Alzheimer's disease are revolutionized by this synergistic synthesis, which makes it possible to identify the disease in its early stages.

The deep learning multi-modal system's functional requirements are the fundamental building pieces that are based on well-defined goals. These requirements, which include feature engineering and data preparation, are meant to build a strong and all-encompassing system that makes use of neuroimaging data. To provide a comprehensive viewpoint on healthcare applications, the word "multi-modal" refers to the integration of many data sources, including genetic data, patient information, and medical imaging. Using deep learning models like CNNs for unstructured data and machine learning approaches like random forests for structured data, the project addresses the difficulties of processing both structured and unstructured data. High-quality, error-free data is ensured through data preparation, which includes feature engineering and thorough data purification. This is necessary for accurate predictions.

Acknowledgements

The author would like to thank Jing Hua Ye and Ruairi O'Reilly for their supervision.

Contents

List of Figures

List of Tables

Abbreviations

AD	**A**lzheimer's **D**isease
ADNI	**A**lzheimer's **D**isease **N**euroimaging **I**nitiative
AI	**A**rtificial **I**ntelligence
AIDS	**A**cquired **I**mmuno **D**eficiency **S**yndrome
ANN	**A**rtificial **N**eural **N**etwork
APOE	**AP**olip**O**protein **E**
API	**A**pplication **P**rogramming **I**nterfaces
Bi-LTSM	**B**idirectional **LTSM**
CAM	**C**lass **A**ctivation **M**aps
CNN	**C**onvolutional **Nl**eural **N**etwork
CT	**C**omputed **T**omography
CPAD	**C**ritical **P**ath **A**lzheimer's **D**isease
CSF	**C**erebro **S**pinal **F**luid
DCNN	**D**eep **CNN**
DL	**D**eep Learning
FDA	**F**ood **D**rug **A**dministration
fMRI	**f**unctional **M**agnetic **R**esonance **I**maging
GA	**G**enetic **A**lgorithm
GAN	**G**enerative **A**dversarial **N**etworks
GBDT	**G**radient-**B**oosted **D**ecision **T**ree
HIV	**H**uman **I**mmunodeficiency **V**irus
HML	**H**ybrid **ML**
IV	**I**ntra **V**enous
LSTM	**L**ong **S**hort-**T**erm **M**emory
ML	**M**achine **L**earning

MRI	**M**agnetic **R**esonance **I**maging
MRS	**M**agnetic **R**esonance **S**pectroscopy
NC	**N**ormal **C**ontrol
NI	**N**euro **I**maging
NLP	**N**atural **L**anguage **P**rocessing
PET	**P**ositron **E**mission **T**omography
RRI	**R-R I**nterval
SVM	**S**upport Vector Machine
RNN	**R**ecurrent Neural Network
PCA	**P**rincipal **C**omponent **A**nalysis
XAI	E**X**plainable **AI**

For/Dedicated to/To my...

Chapter 1

Introduction

This research project's main goal is to use cerebrospinal fluid (CSF) and neuroimaging datasets to predict Alzheimer's disease (AD) in advance, which is an urgent challenge. In order to improve patient outcomes and enable prompt intervention, early detection of AD is essential. As a result, current treatments for AD are often ineffective because the disease is frequently diagnosed when symptoms are more severe. This study attempts to create a predictive model that can recognize minute biomarker alterations suggestive of early-stage AD by utilizing the potential of CSF and neuroimaging data. An all-encompassing strategy to improve the accuracy of early detection is presented by the integration of CSF biomarkers, which offer insightful information about neurochemical changes, and neuroimaging data, which permits precise visualization of structural and functional alterations in the brain.

This research is significant because it has the potential to completely change how AD is diagnosed and treated. Early detection enables focused therapeutic interventions, which may slow the course of the disease and enhance the quality of life for those who are impacted. Combining CSF and neuroimaging datasets allows for the application of sophisticated data analysis techniques and machine learning algorithms, which yields a comprehensive understanding of the complex biological and structural alterations linked to early-stage AD. The project seeks to advance the field of neurodegenerative disease research and pave the way for more effective early interventions in the context of Alzheimer's disease by identifying these early signs and helping to develop reliable and efficient diagnostic tools.

The goal of this project is to create a DL multi-modal predictive model for early AD by integrating neuroimaging and CSF datasets. By extracting complex patterns from both types of data, deep learning within neural networks improves the accuracy of early AD prediction. Carefully selecting and preprocessing CSF and neuroimaging data to make

it compatible with the deep learning model is part of the process. Advanced neural network architectures will be able to understand intricate relationships between CSF biomarkers and neuroimaging features because they can handle multi-modal inputs. The project's goal is to maximize the model's predictive power for early-stage AD detection through rigorous training and validation. This will greatly aid in the creation of precise instruments for clinical intervention and neurodegenerative disease research.

1.1 Motivation

A progressive Disease is a clinical pathological, and/or molecular finding, indicating that the course of a disease is worsening in terms of extent or severity. The benefits of receiving an early AD diagnosis to the patient are that it explains the symptoms and signs they are experiencing as well as putting an end to their suspicions. An early diagnosis and subsequent access to the right services and support can help people take control of their condition, live independently in their own homes for longer, and maintain a good quality of life for themselves, their families, as well as carers. People can plan while they still have the capacity and thus participate in their own legal, financial, and future support/care options and treatment and make their wishes known to the family.[4]

AD is caused by a build-up of brain protein inside and around brain cells, this damages and eventually kills the brain cells. One of these brain proteins is called amyloid, Amyloid can build up and create plaques around brain cells. The other protein is called tau, when tau builds up it can cause tangles inside the brain cells. It is currently unknown how this exact process begins, however the process begins many years before symptoms appear. Over time, as the brain cells die, areas of the brain shrink. The first areas usually affected are responsible for memories known as the hippocampus within the temporal lobe. It is important to pay attention to Alzheimer's as it is currently reaching epidemic levels. AD is seen as the main health and social care challenge of the 21st century as there is currently no cure for the disease.[5]

Clinical trials are completely essential to advancing AD research at a time like this. Through clinical studies conducted over the last 20 years, scientists have made tremendous efforts in understanding how Alzheimer's affects the brain. More than 100 Alzheimer's clinical trials are now recruiting participants. It is only through clinical studies that we will develop and test promising new strategies for treatment, prevention, diagnosis, and ultimately a cure for AD. Unfortunately, clinical trials often suffer from slow or insufficient enrollment, take longer to complete, and are more expensive than trials in most other therapeutic areas. The recruitment and retention of a large number of qualified,

diverse volunteers to participate in clinical research studies remains a challenge to the successful completion of Alzheimer's clinical trials.[6]

As clinical trials can be quite difficult to conduct as well as receive results this drawback greatly enforces the need for an early detection of Alzheimer's. An early diagnosis helps both the person and caregivers learn about Alzheimer's, set realistic expectations, and plan for the future. Many treatable conditions can produce symptoms similar to dementia. For example, vitamin deficiencies, thyroid disease, sleep disorders, alcohol abuse, or depression. Earlier in the disease process one is able to participate more actively in one's own healthcare decisions as opposed to later.[7]

1.2 Contribution

To improve early AD detection using a deep learning multi-modal approach, this project is guided by fundamental research questions. We aim to determine the efficacy of our proposed method by investigating how early AD prediction accuracy is improved by combining CSF and neuroimaging datasets. Simultaneously, we investigate complex links and patterns that are derived from both kinds of data using neural networks, pinpointing particular attributes that are essential for early identification. We will evaluate how careful data curation affects the deep learning model's compatibility with CSF and neuroimaging data. Furthermore, we want to assess how well sophisticated neural network architectures learn intricate correlations between CSF biomarkers and neuroimaging features. These investigations seek to maximize predictive capacities, supporting the creation of precise instruments for early-stage AD identification and promoting clinical intervention and research on neurodegenerative diseases.

We have integrated our deep learning multi-modal approach into a comprehensive system architecture for AD detection. This architecture efficiently integrates neuroimaging and CSF datasets using neural networks. The public dataset will be used to thoroughly assess the system's performance, guaranteeing a reliable evaluation of its predictive abilities for early AD detection. Our methodology consists of training and optimizing the system on various samples from the public dataset, which enables a comprehensive analysis of its precision, sensitivity, and specificity. The selected public dataset serves as a reference point for validation and comparative analysis, allowing us to assess the effectiveness of our suggested system with accepted norms within the scientific community.

1.3 Structure of This Document

Chapter 2 describes the efforts made to date and those being made to use artificial intelligence (AI) to forecast AD. Chapter 3 is a summary of how to predict AD using deep learning (DL). Chapter 4 details the planning and how it will be implemented. Chapter 5 outlines the current conclusion.

Chapter 2

Background

In this chapter, we will review some existing studies conducted in the field of multimodal Alzheimer's detection with an intelligent model.

2.1 Alzheimer's Disease

Alzheimer's is a debilitating irreversible and incurable neurodegenerative disease. AD is caused by a combination of age-related changes in the brain, along with genetic, environmental, and lifestyle factors. AD is characterized by changes in the brain including amyloid plaques and neurofibrillary or tau, tangles that result in a loss of neurons and their connections. This impacts a person's ability to remember, think, and eventually, to live by oneself. Less than 10% of AD patients have early-onset while the majority of cases have late-onset, meaning that symptoms start to show in the middle of people aged 60+. Health, environmental, and lifestyle factors may be a contributor to the disease. Ongoing research will help us understand whether and how reducing risk factors for these conditions may also reduce the risk of AD. AD consists of 3 main stages such as early-stage, middle-stage, and late-stage.

AD can be classified as either early-onset or late-onset. Genetics plays a role in both kinds. The distinctions between AD with late and early onsets are presented in table 2.1

Early-Onset Alzheimer's	**Late-Onset Alzheimer's**
The onset of symptoms usually occurs in the mid-60s to late 30s	Early symptoms start to show up in the mid-60s
Extremely uncommon	Most prevalent kind
Caused mostly by genetic alterations inherited from parents	Potentially involving the apolipoprotein E (APOE) 4 gene

TABLE 2.1: The differences between early-onset and late-onset

AD diagnosis requires a multidisciplinary approach because no single test yields a conclusive diagnosis. Physical examinations, diagnostic tests, like blood and urine analysis, and a comprehensive review of medical history are all part of the process. While cognitive, functional, and behavioral tests gauge thinking and memory, neurological exams evaluate reflexes, coordination, eye movements, and other abilities. CSF tests and brain imaging (CT, MRI, PET) aid in ruling out other illnesses. Genetic testing can reveal risk factors, and mood tests can detect mood disorders that could impair memory. Blood tests show promise in predicting changes in the brain, even though they are not yet FDA-approved. Consistent assessments are provided by digital cognitive tests and devices such as CANTAB and Cognivue. Present-day blood tests are a supplement to a thorough diagnostic workup for AD; they are not-stand-alone diagnostic instruments.

AD is currently treated with the goal of symptom relief rather than brain cell deterioration regression. Although memantine and cholinesterase inhibitors are frequently used, experts are cautiously optimistic about potential future treatments that could alter the course of the disease. Numerous strategies are being investigated in ongoing research, such as using medications, like donanemab and lecanemab, to target beta-amyloid plaques. To reactivate synapses, saracatinib - first developed as a cancer treatment - is being tested for AD. There are difficulties with therapies that block the gamma- and beta-secretase enzymes. Further research is being done on the effects of insulin, sargramostim, tau aggregation inhibitors, and reducing inflammation. The relationship between heart health and AD is investigated, looking at blood pressure treatments and way of life decisions. The role of hormone replacement therapy is still unknown and needs to be investigated further. While it takes time to develop new drugs, initiatives like the critical path for the AD consortium seek to speed up the process through standards and data sharing, providing hope for more potent treatments.

2.2 Data Collections

A basic procedure that entails getting information directly from people or sources using a variety of techniques is manual data collection, often known as elicitation. Surveys are a popular method in which researchers create structured interview questions or questionnaires to elicit specific information from participants. In-person interviews, knowledgeable interviewers can lead candidates through a series of questions and modify the discussion in response to their responses. Observation is an additional technique where researchers watch and document behaviors, events, or activities in a specific setting or systematic basis. Document analysis is another kind of manual data gathering in which researchers examine and extract data from preexisting records, documents, or archive sources. Careful planning is necessary for this process to ensure that observations or questions are objective, transparent, and in line with the goals of the study.

A wide variety of data is gathered in the field of literature to facilitate different kinds of computer analysis and research questions. The main focus is on textual data, which includes written works like novels, articles, essays, and other literary genres. Manual annotation is one way to collect this data, wherein subject matter experts tag or label particular textual qualities so that ML algorithms can identify trends and connections. Text mining methods, including NLP, are used to glean important themes, attitudes, or insights from massive amounts of textual data. Literature data may also contain metadata which is useful for bibliometric research and for comprehending the context of scholarly works. Examples of this metadata include publication dates, author details, and citation counts. The collection of data for computing comprises several methods. One popular technique is web scraping, in which computer programs automatically retrieve pertinent data from websites, databases, or digital files. API offers a standardized method for accessing and retrieving data from several platforms, making it easier to integrate data into computation workflows. Additionally, user-generated content, reviews, and discussions on social media and collaboration platforms increase the amount of data available and present academics with options for data mining and analysis. Ultimately, a diverse method for gathering and harvesting literary material for computation needs is present by the mix of manual annotation, text mining, online scraping, and API use.

Data from imaging techniques that record the anatomical or functional features of the brain are referred to as neuroimaging data. MRI, fMRI, PET, CT are common neuroimaging modalities. These methods produce finely detailed pictures of the blood flow, metabolism, and structure of the brain. Utilizing specialized equipment to take non-invasive brain scans is part of the process of gathering neuroimaging data. While fMRI uses changes in blood flow to identify brain areas linked to particular tasks or stimuli,

MRI uses strong magnets and radio waves to provide detailed structural images. A radioactive tracer is injected during PET to see metabolic activity. Different features are used in neuroimaging analysis to extract information about the anatomy and function of the brain. Measures of the brain's surface area, thickness, and volume are examples of structural traits that are frequently employed in research on neurodegenerative illnesses and developmental abnormalities. To comprehend cognitive processes or responses to stimuli, function characteristics entail patterns of brain activity and connections that are evaluated by fMRI. By analyzing brain activity at the voxel level, voxel-based analysis offers spatial insights. Furthermore, region-of-interest analysis concentrates on particular brain areas. While ML approaches may use a variety of features for tasks such as classification or prediction, graph theory is utilized to analyze patterns of connectedness. In general, a wide range of anatomical and functional characteristics are covered by neuroimaging data analysis, which makes it possible to fully comprehend the complexities of the brain.

The transparent fluid that surrounds the brain and spinal cord that is essential for maintaining and nourishing the central nervous system is referred to as CSF data. A spinal tap or lumbar puncture involves inserting a needle into the spinal canal to remove a small amount of fluid, which is how CSF is obtained. Healthcare personnel usually carry out this operation in a clinical setting. The diagnosis of a number of neurological problems, such as infections, inflammatory disease, and neurodegenerative illnesses like AD, is made possible by CSF examination. Among the characteristics that are frequently used for CSF analysis are biochemical markers, which provide information about the fluid's composition. Certain proteins-like tau and beta-amyloid - are essential to understanding AD. Potential neurodegenerative processes may be indicated by the presence and concentrations of these proteins in CSF. Furthermore, measurements of total protein content, glucose concentrations, and cell count are regularly checked. More sophisticated methods, including mass spectrometry, can examine the molecular makeup of CSF in more detail. When these characteristics are combined, researchers and medical experts can evaluate the CSF's biochemical profile and obtain important data for neurology-related research and diagnostics.

The established research dataset in this field is presented in table 2.2.

Database	**Modalities**	**Stages**
ADNI	Neuroimaging + APOE + CSF	CN, SMC, MCI

Table 2.2: Dataset for Alzheimer's Detection

2.3 Hyperparameter Optimization and Attention Mechanism

2.3.1 Hyperparameter Optimization

A crucial step in building ML models is hyperparameter optimization, which entails fine-tuning parameters that were not discovered during training. The Grid-Search approach [8] is a widely used method for Alzheimer's detection that has been mentioned extensively in research papers. Grid-Search involves methodically examining a predetermined range of hyperparameter values and assessing model performance for each combination in order to determine which is the best. Grid-Search, which is highly regarded for its transparency and simplicity, is used to optimize model performance by fine-tuning algorithms in Alzheimer's research, especially in the analysis of neuroimaging and related datasets.

GA [9] are extensively used in hyperparameter tuning for the identification of AD. The literature has made use of some GA variations, such as GA with elitism, tournament selection, crossover, and mutation. Motivated by natural selection, GA efficiently explores a variety of search spaces by iteratively evolving a population of hyperparameter sets. GA's ability to adjust to intricate hyperparameter setups makes them a good choice for optimizing ML models. Their capacity to effectively converge towards optimal or nearly optimal solutions enhances the efficacy of the model in detecting AD. In general, GAs are essential for optimizing hyperparameters, which improves the precision and dependability of models used with Alzheimer's datasets.

In ML, the attention mechanism is a key idea that improves models' ability to concentrate on pertinent information while analyzing data. The method selectively applies weights to input elements, enabling the model to allocate different weights to different segments of the input sequence. Transformer-based architectures and self-attention [10, 11] are two methods with which attention mechanisms are frequently employed. A model that uses self-attention can take into account the relationships between each element in a sequence at the same time. Transformer architectures make use of multi-head attention to capture various facets of data interdependency. To allow the model to dynamically highlight important features during processing, attention mechanisms are embedded into models by adding attention layers or modules. Because of their versatility, attention mechanisms are a useful tool in many ML applications as they improve the model's capacity to extract meaningful patterns and relationships.

Model performance is improved when hyperparameter tuning and attention mechanisms are combined. By optimizing parameters that are crucial for attention mechanisms, hyperparameter optimization—such as that achieved through the use of genetic algorithms—fine-tunes the overall configuration of the model. During data processing, attention mechanisms like Self-Attention concentrate on pertinent features. Optimizing hyperparameters related to attention layers, such as the quantity of attention heads or learning rates, is necessary to integrate the two. By addressing the nuances of both model architecture and attention mechanism parameters, this dual optimization improves the synergy between the two [12]. The model's ability to detect complex patterns and allocate attention efficiently is ensured by the iterative refinement of hyperparameters and attention mechanisms. This leads to improved performance, especially in difficult tasks like Alzheimer's detection where it is essential to fine-tune these components to achieve the best results.

James Hall et al [13] consider the expanding datasets containing blood and MRI biomarkers, which highlights the need for accelerated ML in AD diagnosis. Specifically, a multicore high-performance workflow for hyperparameter tuning of SVM is studied to improve computational efficiency. The model is shown using publicly available MRI data, using 100 iterations of 5-fold cross-validation and factoring in demographics. Findings demonstrate the high-performance hyperparameter tuning model's potential for accelerating ML in AD diagnosis, with a noteworthy 96% increase in computational efficiency. A further indication of the model's wider applicability in advancing AD biomarker applications is its adaptability to different ML algorithms, including logistic regression, random forest, and XGboost.

To fine-tune the hyperparameters for AD prediction, Fan Zhang et al [14] created a very high-performance computing process utilizing parallel SVM. They demonstrated a 96% reduction in computational time and highlighted the crucial role that MMSE, Age, Sex, nWBV, and SES play in AD diagnosis. Melissa Petersen et al [15] improved efficiency and accuracy in predicting AD and mild cognitive impairment (MCI) in imbalanced data from the HABS-HD project by using high-performance computing to produce a 98.2% reduction in computational time for SVM hyperparameter tweaking. Using ADNI-standardised MRI datasets, S. Raja et al achieved improved accuracy in hippocampus segmentation and AD classification compared to other classifiers by using Deer Hunting Optimisation [16] to optimize hyperparameters in a Capsule Network model. With 86.84% accuracy using a Random Forest classifier, Afreen Khan et al created a strong ML pipeline [17] for dementia prediction using longitudinal MRI data, highlighting the need for more advancements in AD diagnosis and treatment.

The study conducted by Kandasamy et al showcased enhanced early AD detection performance compared to baseline classifiers by the optimization of hyperparameters and the use of SMOTE [18] in an ensemble-based model. The findings also suggested potential improvements and clinical applications. Through ML, Monika Sethi et al [19] improve the early identification of AD by fine-tuning hyperparameters of CNN frameworks using Bayesian search and showcasing the effectiveness of Bayesian optimization for better model performance. Aleena Thomas investigates the detection of AD using various CNN models, highlighting the significance of striking a balance between accuracy and precision. She highlights LeNet's [20] impressive precision of 94.31% and points out that improved performance is not guaranteed by model complexity alone, indicating that future research should concentrate on optimization strategies and hyperparameter refinement for larger datasets.

Using non-linear neural networks and XAI techniques, Abraham Varghese et al. [21] increase interpretability in AD classification, providing transparent models with meaningful explanations for medical practitioners in diagnosis and treatment decisions. With attention-based cross-modal interactions, Michal Golovanevsky et al [13] presents a multimodal DL framework for AD detection that achieves greater accuracy (96.88%) and suggests its possible integration into clinical settings for accurate disease diagnosis.

2.3.2 Attention Mechanism

Rude Lin et al presented Brain Informer (BraInf) [22], a cutting-edge DL model for effective feature encoding and classification of MRI data in AD. BraInf uses a self-attention mechanism. BraInf uses a multi-head ProbSparse self-attention mechanism to overcome the limitations of the self-attention mechanism, which lowers computational complexity and allows the mechanism to be applied to high-dimensional MRI data. Furthermore, a structural distilling layer minimizes memory costs while guaranteeing network depth. BraInf outperformed state-of-the-art techniques with superior classification accuracies of 97.97% for NC/AD and 91.89% for MCI/AD when tested on the ADNI dataset. BraInf's effective representation of learning ability in brain structural DL data is demonstrated by ablation studies, which provide encouraging results for DL applications in brain disease research.

Hyeon Kang et al [23] showed that dual-phase FBB imaging, in conjunction with LSTM and attention mechanisms, produces higher accuracy in AD positivity scores and stronger correlations with psychological tests compared to delay-phase FBB imaging alone. They achieved this by using an attention mechanism to improve connectivity in modular networks. With intentions to address constraints through dataset extension, XAI, and

clinical deployment for increased clinical applicability and interpretability, Shui-Hua Wang et al presents ADVIAN [24], an AD VGG-Inspired Attention Network, which is effective in AD diagnosis. With high accuracies (97.35%, 87.82%, 78.79%) on the ADNI database for AD detection and MCI conversion prediction from structural MRI data, Jie Zhang et al introduce CAM-CNN [**?**], a densely connected CNN with a connection-wise attention mechanism.

With the potential for wider applications in clinical diagnosis and early prevention, Xiaofei Sun et al [25] presents a novel attention-based learning framework that integrates MRI and EHR data for precise brain degeneration diagnosis. By leveraging intramodal and intermodal dependencies through spatial-temporal and cross-attention mechanisms, the authors achieve superior performance in prediction. Shuihua Wang et al introduce WS-AMN [26], a DL model that combines weak supervision learning with attention mechanisms. This model achieves exceptional accuracy (99.61%) in the classification of AD on the OASIS dataset, overcoming the difficulties presented by small AD datasets for accurate disease identification in the medical domain. A DCNN with multiple attention mechanisms, cyclic convolution for feature stability, and an improved VGG backbone model are used by Liu Fei et al [27] to predict AD with a 99.8% accuracy rate. This is a novel method for precision diagnosis through MRI integration.

Using MRI on the OASIS dataset, Faiez Gargouri et al propose a Depthwise Separable Convolutional ResNet with attention mechanisms [28], which they show to be effective in multimodal binary classification with a 92% accuracy rate and achieve 99% accuracy in AD diagnosis. In the ADNI dataset, Athena George et al [29] used a 3D CNN with channel and spatial attention mechanisms on MRI volumetric scans, and they achieved an impressive 87% accuracy for AD detection. This shows the promise of DL in AD classification, though more real-world testing is necessary. By identifying important brain regions and attaining enhanced classification performance, Dan Jin et al propose an attention-based 3D ResNet [30] for AD diagnosis using neuroimaging, demonstrating the promise of DL in biomarker discovery and disease classification. Yin Chen shows the effectiveness of a dual attention mechanism spatio-temporal graph CNN[31] by using a CNN with attention mechanisms, such as feature-level and regional attention, to obtain superior AD classification in brain MRI data.

2.4 Modality

The literature suggests that structural MRI is a prominent single modality for the detection of AD, although multimodal approaches are frequently employed. The brain's anatomy can be seen in great detail thanks to structural MRI [32], which can also identify

AD-related atrophy patterns. Methods like cortical thickness measurement, volumetric analysis, and evaluating the volumes of particular regions are frequently used. Early disruptions typical of AD are revealed by diffusion tensor imaging, which assesses white matter integrity. Combining ML algorithms—like neural networks and SVM — improves diagnostic precision by using big datasets to identify intricate patterns. As multimodal strategies are common, a thorough structural MRI analysis using a variety of techniques offers important insights for Alzheimer's detection.

Among the fundamental single modalities in neuroimaging, MRI stands out for providing comprehensive structural and functional insights into the brain. The high-resolution images of the brain anatomy provided by structural MRI help identify abnormalities or lesions. Changes in blood flow during tasks or at rest are measured by fMRI. By following the motion of water molecules, diffusion MRI can clarify white matter tracts and connectivity patterns. Using MRS, one can measure biochemical markers and learn about the concentrations of metabolites in different brain regions. Intrinsic connectivity networks are revealed by sophisticated methods such as resting-state fMRI [33]. All of these MRI-based methods work together to provide a thorough understanding of the brain's structure and function, making MRI a flexible and essential tool for neuroimaging.

Multimodal neuroimaging combines multiple methods to provide a thorough examination of the anatomy and physiology of the brain. Combining fMRI to record brain activity in real-time with MRI to provide detailed anatomy is a common method. Metabolic processes are frequently revealed through the integration of PET. This combination allows for a more comprehensive understanding of neurological disorders by allowing for a nuanced correlation to be seen between structural anomalies and functional alterations. By enhancing the simultaneous analysis of structural and functional data, methods such as voxel-based morphometry (VBM) [34] improve localization accuracy. Understanding the complex relationship between brain structure and function through multimodal neuroimaging is crucial to improving our understanding of neurological disorders.

Basit Raza et al [35] investigates the improvement of brain parcellation-based computer-aided approaches for the identification of AD and MCI, using various anatomical at lases (LONI Probabilistic Brain Atlas and Automated Anatomical Labelling). By merging characteristics from multiple atlases, the method produces a more thorough and reliable depiction of brain regions. The work uses structural MRI and FDG-PET scans to compute region-specific measures, utilizing data from the ADNI database. Individual and combined characteristics are used to classify AD, MCI, and cognitively normal people. This shows that a multi-atlas strategy enhances classification accuracy over single-atlas methods. Feature ranking illustrates how important particular brain regions

are for precise subject classification and how different atlases can be for more in-depth neuroimaging studies.

The assessment of neuroimaging modalities for the detection of AD by Morteza Amini et al emphasizes the superiority of CNN-based algorithms [36], emphasizes the need to address constraints in data collection, advocates for diverse datasets, and emphasizes the inclusion of different stages of the disease for improved early detection. Using multimodal neuroimaging data, Modupe Odusami et al [37] explore the potential of ML for classifying different phases of AD. Their promising accuracy is highlighted, but further study is required to completely establish its diagnostic potential. Weichen Huang uses multimodal contrastive learning and tabular attention modules [38] to provide a strong framework for detecting AD. It achieves an accuracy of over 83.8% and highlights the significance of particular biomarkers for early identification and monitoring.

A new dataset (VBSD) is introduced for evaluation, and the LogisticRegressionCV model performs optimally. Haibao Chen et al present a spectrogram-based technique [39] for AD identification using voice data, proving feasibility in early detection. Elisa Tuzzi et al [40] uses SVM classifiers on diffusion and structural MRI data, discovering that the "T1+DWI" features when combined provide the highest accuracy in diagnosing AD, surpassing the results of using separate modalities. In their investigation into the identification of AD by speech analysis, Zihao Wu et al [41] make use of pre-trained language models and Graph Neural Networks on the DementiaBank Pitt database. They also use data augmentation and a unique fusion approach for multi-modal analysis. Eric Westman et al [42] compared the classification performance of MRI and CSF biomarkers separately and in combination using a novel technique (OPLS). They found that the combined approach was more accurate in identifying MCI subjects, differentiating AD from healthy controls, and predicting the conversion of MCI to AD. When assessing predictive biomarkers for AD-type dementia in patients with MCI, Stephanie Vos et al [43] discovered that the CSF Aβ1-42/tau ratio performs better than hippocampus volume. They suggest more research into combining CSF and MRI biomarkers with PIB-PET and FDG-PET to improve prediction. James Glass et al apply dimensionality reduction techniques [44], revealing superior performance of text models, and highlight challenges in predicting MMSE scores for mild dementia, suggesting implications for larger-scale studies and possible investigation of correlations with apathetic symptoms in future research. A balanced subset of the Pitt corpus is preprocessed for audio recordings and transcripts, with normalized chunks and voice activity detection.

2.4.1 Multimodality

Yubraj Gupta et al investigate multimodal biomarkers that combine structural MRI, FDG-PET, CSF, and APOE genotype features for the early prediction of AD. Features from various modalities are concatenated using an early fusion technique, necessitating unique combination rules for successful integration (see fig 2.1). Random tree embedding is used to convert low-dimensional features APOE and CSF into a higher-dimensional state to address dimensionality disparities. The feature selection is improved by truncated SVD dimensionality reduction. The suggested approach distinguishes between subjects with MCI who are stable and those who are converters with 94.86% accuracy, 93.59% AUC, and a Cohen's kappa index of 0.86. The suggested method outperforms cutting-edge methods with 158 ADNI dataset subjects, demonstrating its dependability in early AD prediction through the use of an extensive collection of multimodal biomarkers.

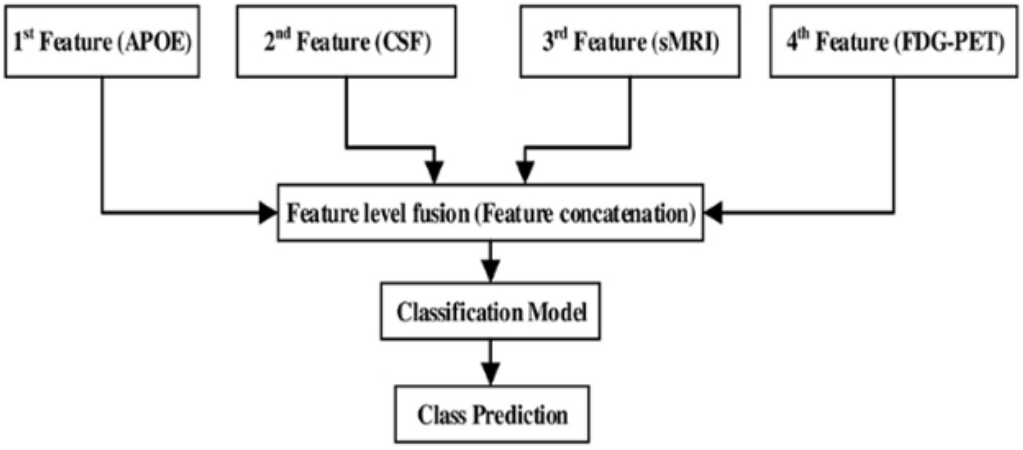

FIGURE 2.1: Feature Selection for AD detection [1]

The use of imaging biomarkers [45] in neurodegenerative disorders is reviewed by Emma Coomans et al. They emphasize the need for ML for early identification and the necessity of integrated techniques and longitudinal investigations in clinical trials to improve diagnostic skills. With potential applications in clinical trials and targeted interventions, Garam Lee et al introduces a multimodal RNN[46] that combines neuroimaging, CSF biomarkers, cognitive performance data, and demographics. Their findings demonstrate enhanced prediction accuracy for the transition from MCI to AD compared to single-modality models. With potential applications in long-term neurological condition research, Armando Barreto et al present a novel multimodal-multitask model [47] for

predicting the progression of AD. The model integrates a variety of biomarkers and optimizes accuracy through the use of modality-specific multitask coefficients and ensemble learning.

To predict neuropsychological scores, Seyed Hani Hojjati et al use multimodal neuroimaging and ANNs [48]. This shows that average fluorodeoxyglucose-PET has a superior predictive power in particular brain regions and that the bimodal approach outperforms unimodal methods in both MCI and AD stages. Additionally, the bimodal approach sheds light on the non-linear progression trend of AD. The relationship between structural MRI changes and amyloid pathology is revealed by Enrico Peira et al [49] using an automated classifier that integrates demographic, cognitive, MRI, and APOE $\epsilon 4$ data. They show that this classifier is useful for pre-screening clinical trial participants for anti-amyloid therapy, with AUCs of 0.81 for MCI and 0.74 for normal cognition. In comparison to a model using only CSF biomarkers, Michael F. Romano et al's fusion model of T1-weighted MRI, FDG PET, and amyloid PET images uses DL to predict 2-year AD progression risk in individuals with MCI with superior accuracy and sensitivity [50]. This highlights potential applications for risk prediction, especially with FDG data.

Uttam Khatri et al use a thorough approach that combines structural MRI measurements and non-imaging biomarkers [51], improving the accuracy of AD diagnosis, especially in difficult classifications. Further research is needed to determine the efficacy of the method and comprehend the progression of the disease, but it may lead to improved AD scanning protocols. Incorporating neurodegeneration markers, Aβ burden, APOE4 status, and advanced MCI stage, Seung Joo Kim et al developed a multimodal biomarker predictive model [52] for fast decliners in Aβ+MCI patients. The model demonstrates exceptional predictive performance with nomograms, providing a useful tool for clinicians in identifying early disease progression. Using the TADPOLE dataset from the ADNI, Thushara et al [53] apply the random forest classification method and achieve accuracy comparable to current research in multiclass classification for the early diagnosis and prediction of AD.

2.5 Models

2.5.1 Machine Learning Model

Rashmi Kumari et al present a multimodal method for early AD detection that combines cognitive testing, metabolic decrease using FDG-PET, amyloid plaque accumulation using PiB-PET, and structural atrophy using MRI. It suggests automated hyperparameter

tuning for effective feature selection and random forest optimization with the adaptive hyperparameter tuning random forest ensemble classifier (HPT-RFE) [54]. In terms of computing speed, the classifier performs better than SVM, Naïve Bayes, K-Nearest Neighbour, and ANN. HPT-RFE outperforms state-of-the-art methods in simulations using the ADNI dataset, achieving 100% accuracy, sensitivity, and specificity for AD against NC and competitive performance for NC versus MCI and AD versus MCI classifications.

Karol Estrada et al [55] evaluated the effectiveness of several feature selection techniques and classification algorithms for predicting AD using cross-validation and ranking based on 95% confidence intervals. Metrics like accuracy, precision, sensitivity, balanced error rate, and ROC AUC in the ADNI dataset were used, with validation on a separate subset. They emphasized a classification performance of roughly 72% area under the ROC curve, ML algorithms are promising in terms of predicting genetic risk for late-onset AD (LOAD). They also provided a means of finding new genetic markers that may be linked to the condition. Ezzati Ali et al introduce ensemble linear discriminant models [56] that use volumetric and demographic MRI features to predict the clinical progression of participants with MCI with notable accuracy over time. These models show high accuracy in differentiating AD from cognitively normal individuals. Sinead Gaubert et al highlight EEG as a portable tool for neurodegeneration prediction and highlight the possibility of ML in early Alzheimer's risk assessment utilizing accessible biomarkers [57]. However, difficulties remain in correctly diagnosing preclinical AD in people with subjective cognitive impairment. To improve predictive accuracy, future research should incorporate varied pre-clinical cohorts, use longitudinal multimodal measures, and make use of polygenic risk scores.

2.5.2 Deep Learning Model

Scott A. Small et al [58] examine the pathologies associated with AD and suggest a voxel-based DL model that uses sMRI to predict neurodegeneration more accurately than other neuroimaging techniques. After using data augmentation for training, the model performs better than amyloid and tau biomarkers in prodromal AD (AUROC = 0.788) and correctly detects Alzheimer's dementia (AUROC = 0.973). Specifically, for tau pathology (CSF tau = 0.682) and amyloid pathology (CSF Aβ = 0.702), the DL model outperforms CSF biomarkers. The results demonstrate the model's effectiveness in locating biomarkers associated with AD from traditional MRIs, pointing to possible advantages in terms of lowering patient burden, risk, and expense.

Andrea Loddo et al present a deep-ensemble method [59] that compares different DL models across three MRI and one fMRI dataset to diagnose AD from brain images. Achieving automated dementia-level categorization, choosing the best architecture, and testing robustness are among the goals. This method outperforms the state-of-the-art in both binary (98.51% accuracy) and multiclass (98.67% accuracy) classifications on a variety of datasets. The suggested approach shows promise for stable and dependable computer-aided diagnosis systems, as it is intended to support clinical care based on brain scans. Its adaptability to other imaging techniques is presented, highlighting its potential benefits for patient care, including successful tests on MRIs and fMRIs.

Abdulaziz Alorf et al classify six phases of AD using resting-state functional MRI (rs-fMRI) data by applying DL algorithms, namely Stacked Sparse Auto-Encoder Network (SSAE) and Brain Connectivity-based Graph Convolutional Network (BC-GCN) [60]. Functional connectivity matrices may be extracted using a thorough preprocessing pipeline, and SSAE and BC-GCN are optimized for binary and multi-label classification. CN vs. SMC and the six-stage classification with BC-GCN yield high accuracy rates of 92.75% and 84.03%, respectively, for the suggested techniques. The frontal gyrus and precentral gyrus are two important brain regions identified by the study. SSAE and BC-GCN offer a useful tool for AD diagnosis and biomarker identification, outperforming earlier AD classification techniques on rs-fMRI data.

M. Baskar et al use multi-class log loss and MRI scan analysis [61], a hybrid DL model for AD stage identification that shows better accuracy and lower prediction error than previous methods. Kyung-Ah Sohn et al [62] demonstrates the potential for risk identification and stratification in clinical trials by introducing a multimodal RNN that uses DL and integrates demographic, longitudinal CSF biomarkers, cognitive performance, and cross-sectional neuroimaging data to predict the conversion of MCI to AD with improved accuracy. Hongyoon Choi et al demonstrate superior accuracy (84.2%) in predicting the conversion of MCI to AD compared to traditional feature-based quantification approaches, showcasing the potential of DL as a useful tool for predictive neuroimaging biomarkers, introducing a DCNN using fluorodeoxyglucose and florbetapir PET images [63]. Krish Desai et al use gradient boosting decision trees in conjunction with CNNs (ResNet, EfficientNet, and RegNet) [64] to predict the amyloid standardized uptake value ratio (SUVR) from PET scans. The results show 96.4% accuracy on a test set, suggesting a potentially effective and automated method for SUVR calculations in AD that can be accessed through the DeepAD web application.

2.5.3 Hybrid Models

2.5.4 Hybrid Machine Learning Models

Hybrid ML (HML) addresses the shortcomings of individual approaches by combining multiple algorithms, procedures, or strategies from distinct domains to improve overall performance. Before employing HML algorithms, data pretreatment techniques like principal component analysis (PCA) or linear correlation analysis are frequently used. These techniques provide a special workflow that is different from that of typical ML. HML emphasizes the necessity of combining complementary strategies since it acknowledges that a single ML methodology might not be appropriate for every task. HML seeks to develop hybrid models that perform better than individual techniques by utilizing a variety of approaches. This methodology offers a flexible and efficient means of tackling difficult problems across various domains and data sources.

Ahmet Cinar et al [65] emphasized the use of an enhanced ResNet50 model in a computer-aided system for the early detection of AD. Through layer adjustments, the ResNet50 architecture was improved to the point where the accuracy rate went from 78% to 90%, exceeding previous research. This study's high-performance methodology was demonstrated by the use of a dataset that represented the four stages of the disease, in contrast to numerous two-class examinations. AD diagnosis may advance thanks to the hybrid model that has been built, which shows encouraging findings with a 90% accuracy rate. Early detection is critical in AD.

Gopi Battineni et al [66] used a longitudinal dataset of 373 MRI sessions from 150 participants to create ML models for the early detection of AD and the prediction of dementia in the elderly. The hybrid model improved accuracy significantly to 98% by combining 1NN and SVM with human and automatic feature selection. The study highlights the difficulties in detecting AD in its early stages, the potential of ML to predict the risk of dementia, and the necessity of taking into account both genetic and non-genetic components in future research.

Kamna et al [67] present a thorough strategy for predicting the course of AD, including data exploration, preprocessing, and a hybrid model that combines algorithms from decision trees and logistic regression. The hybrid model outperforms other models with an astounding 96% accuracy rate after extensive testing. This discovery has promise for early intervention and optimal treatment techniques, and it makes a substantial contribution to the field of Alzheimer's prediction. The efficaciousness of the suggested methodology highlights its capacity to improve diagnostic accuracy and steer treatment approaches, thereby tackling significant obstacles in the management of AD.

H. Azath et al [68] utilize CNN methods to improve AD diagnosis accuracy by 20% and highlight the efficacy of Hybrid AI, which combines CNN and SVM to forecast AD stages based on MRI data. Swaleha Zubair et al [69] used cognitive and demographic data, a complex 3-tiered ML algorithm that outperforms other classifiers through model stacking to achieve up to 95.12% accuracy in predicting AD and MCI, providing useful early diagnosis and validation across ADNI datasets. Babu et al uses MRI images to present a novel predictive approach for early AD detection. It incorporates optimal feature selection using the CG-DU hybrid model [70], feature extraction using GLCM and Haralick techniques, and CNN and SVM combination classification, showing improved performance over previous methods.

2.5.5 Hybrid Deep Learning Models

To improve overall performance, a hybrid DL model is an integrated computational framework that integrates different DL architectures or incorporates non-DL components. It optimizes learning from a variety of data sources and makes use of the advantages of several models to handle certain problems. In the hybrid approach, conventional ML methods are frequently combined with CNNs, RNNs, or other DL architectures. By utilizing each component's strengths in concert, this synergistic integration attempts to create a more resilient and adaptable model. The efficacy of hybrid DL models in collecting intricate patterns and enhancing predictive accuracy has been demonstrated through their successful use in several areas, including medical diagnosis, image processing, and natural language comprehension [71].

Jose M. Alonso et al [72] focus on the vital issue of providing elderly people with reliable progression predictions for AD. MRBL and DFBL, two new hybrid DL models, are presented and assessed with the use of ADNI data. A clinical decision support system is made possible by MRBL, an interpretable multitask regression model that predicts seven cognitive scores 2.5 years after observation. DFBL is a hybrid model that trains several classifiers by utilizing deep features from bidirectional SVM (BiLSTM) models. The models' effectiveness is evaluated on 1371 people using a variety of modalities, highlighting their applicability in real-world situations.

Hager Saleh et al provide a strong hybrid CNN-LSTM model [73] that can be used to predict the course of AD. It combines four cognitive sub-score modalities with GAs and Bayesian optimization to enhance explainability. This model performs better on the ADNI dataset and may be used to support decisions in the real world. Sara Buttau et al provide a potent deep-ensemble method [59] for classifying AD from a variety of brain pictures. This method achieves remarkable accuracy and shows promise for

reliable computer-aided diagnosis in clinical settings. Nivedhitha Mahendran et al [74] use ML classifiers, the current hybrid gene selection pipeline—which combines mRmR, WPSO, and AE—performs better than previous approaches in mitigating the "curse of dimensionality" in molecular data. It is planned to be evaluated on SNP and DNA Methylation datasets in the future.

Shujuan Liu et al present a hybrid DL framework [75] that combines neuropsychiatric symptoms, age, gender, behavioral scores, and structural MRI to provide superior diagnostic accuracy for AD. P Dheepan et al [76] analyze ventricular deformation in AD using transverse MR images using automated segmentation with U-Net and SegUnet. ResNet-101 is used for feature extraction, and a classifier merging strategy is used for efficient classification and comparison.

2.6 Time Series

A time series represents the evolution of a particular variable over a specified period through a sequential collection of data points collected at regular intervals. On the other hand, data that is cross-sectional is information that has been collected all at once. Time series analysis is a tool used in the investment world to monitor and analyze data point movements over time, such as security prices. On the other hand, cross-sectional analysis looks at data at a particular point in time and helps investors find the companies that are doing the best out of a given set of metrics. Investors frequently use a combination of these methods to evaluate past performance (time series) and cross-sectional comparisons with peers or industry benchmarks to make well-informed decisions. The focus of cross-sectional and time series data, which are frequently combined in practical analysis, is on a single point in time or the evolution of data over time, respectively.

P.J. Moore et al present a novel method [77] that integrates time series analysis into a DL framework for predicting the progression of AD disease. The study uses a random forest model in recognition of the difficulties caused by missing and randomly sampled data points in time-dependent studies of AD. With the help of an input vector that contains demographic and non-time-varying variables, like genetic data, and summarises the history of the time series, this model is specifically made to identify relationships between pairs of data points at various time intervals. During the experimentation phase, the model is tested using data from the TADPOLE grand challenge, a project that uses demographic, physical, and cognitive input data to predict the evolution of people at risk of AD disease. Predicting the diagnosis, ADAS-13 score, and normalized ventricle volume are among the tasks. The study is noteworthy for contrasting its approach with a benchmark SVM predictor, exhibiting better results with a mAUC of 0.82 and

a classification accuracy of 0.73, as opposed to the SVM's mAUC of 0.62 and BCA of 0.52 for diagnosis prediction.

J. Kim employed a six-layer multi-perceptron DL classification model [78], 78 senior citizens' dataset was examined to identify patients with MCI associated with AD within the normal group. Two hundred and twenty training sets and two hundred test sets were produced by randomly assigning 79 patients to ten groups, all with a patient ratio of normal to MCI. An Intel Core i7-12700F 4.9 GHz processor, 512 GB RAM, and an NVIDIA GEForce RTX 3080 TI GPU were used to develop the model, and Optuna was used for hyperparameter tuning. Several software tools were used in the analysis, including sci-kit Learn, TensorFlow-GPU, Python, Keras, NumPy, Pandas, and Matplotlib. Features with a two-sided p-value in the t-test less than 0.1 were chosen for ML. One-dimensional representative values from time series data were compared between the groups with mild and normal cognitive impairment. The findings showed that in 120 DL models, 150 features that were taken from time series data had an accuracy range of 0.78-0.90. The models' respective ranges for sensitivity, specificity, recall, precision, and F1 score were 0.88-0.96, 0.86-0.94, 0.78-0.90, 0.78-0.90, and 0.74-0.88. The performance of the model was shown by the area under the curve (AUC), which was computed from the receiver operating characteristic curve. Using comprehensive features from functional near-infrared spectroscopy (fNIRS) time-series data, a novel DL algorithm with high sensitivity. Particularly, significant features were determined to be Hjorth, Kurtosis, skewness, entropy, curve length, AUC, autocorrelation, and time to peak. According to the authors, this algorithm may help identify people who have MCI from AD from those who do not.

Nasir Rahim et al present a predicting framework that uses a hybrid 3D CNN and bidirectional RNN (3D-CNN-BNN) [79] to predict the progression of AD. This model outperforms earlier models by integrating feed-forward neural networks, RNNs, and 3D CNN. The model predicts a patient's condition three years later using longitudinal 3D MRI volumes. With a BRNN module created especially for inter-volumetric relationships, the end-to-end architecture captures both intra- and inter-slice features. The integration of cognitive and demographic biomarkers with the model is investigated in this work. Furthermore, by employing guided Grad-CAM to track brain tissue regions over time, the study pioneers time-series visual explainability of 3D MRI neuroimages for AD, supporting the model's conclusions. The effects of adding multimodal input data were investigated in four experiments: MRI scans without any demographic features, MRI scans with demographic features, MRI scans with clinical scores (CSs), and MRI scans with demographics and CSs for progression detection. The study also looked at how the accuracy of the model was affected by increasing time steps in the input data. To ensure appropriate training and testing stratified 5-fold cross-validation was used.

Accuracy, precision, recall, and area under the curve (AUC) were the four evaluation metrics used to determine how well each model performed on average.

Shaker El-Sappagh et al present a new sophisticated multimodal multitask DL architecture [80] that is presented in this work to detect the progression of AD. Comparing the proposed framework to current state-of-the-art studies, a more stable and accurate system is produced by jointly predicting numerous variables using time series data from various sources. Using a multitask method, the model predicts a patient's advancement status as well as four important cognitive scores simultaneously. In contrast to early research, this model builds on static baseline features as background knowledge and extracts temporal features from five different types of heterogeneous data sources to achieve better performance. The effectiveness of the suggested model is validated by extensive tests conducted on a dataset of 1536 patient samples from the ADNI database. These tests highlight the significance of statistical features extracted from time series data, the superiority of deep-stacked CNN-BiLTSM, and late feature fusion.

T.J. Lyons et al [77] tackle the difficulties associated with classifying AD by evaluating an approach's performance using data from the TADPOLE grand challenge and contrasting the outcomes with guidelines supplied by the competition's organizers. The article highlights approaches and best-performing models while discussing earlier issues like CADDementia and Kaggle Neuroimaging. The TADPOLE grand challenge, which is still in progress as this research is published, consists of a three-fold classification and the use of various measurement modes to predict the ADAS-13 score and normalized brain volume. The ADNI-1, ADNI-GO, ADNI-2, and ADNI-3 phases of the ADNI are the sources of the dataset. The purpose of the study is to forecast future data obtained from the ADNI-3 phase and provide specifics on the TADPOLE competition structure, including leaderboard datasets (LB1, LB2, LB4). The outcomes, which are based on the leaderboard dataset, show how difficult it is to forecast AD, MCI, and control groups with any degree of accuracy. There are some significant errors in the diagnosis of MCI and AD. The study shed light on the complexities of predicting the progression of AD and the evaluation metrics involved, emphasizing how difficult it is to predict changes between diagnostic stages over a long period.

Tamer Abuhmed et al provide a hybrid DL architecture [81] that integrated bidirectional LSTM (Bi-LTSM) and a lightweight DCNN for the identification of AD development. Using temporal rank pooling, the DCNN creates 2D dynamic pictures from longitudinal 3D MRI volumes, capturing the anatomy of the brain. A CNN-Bi-LTSM model processes these images and patient cognitive scores to predict the progression of Ad at 48

months, providing a multimodal approach for improved accuracy. To help doctors monitor changes over time, the study presents a unique XAI technique for visualizing temporal features. Comprehensive experiments on the ADNI dataset show better performance than known DL architectures. Additionally, using guided GRAD-CAM to examine the interpretability of model decisions, the study identifies key brain regions. Future studies will look into learnable temporal rank pool algorithms and different modalities like PET in an effort to offer thorough and comprehensible insights for AD diagnosis and comprehension.

Farman Ali et al present an ensemble learning framework [82] based on large multimodal time series data; this work makes a substantial addition to the identification of AD development. The ensemble classifier's stability and accuracy are improved by the framework's integration of early data fusion and late decision fusion features. Four medically important modalities with four-time timesteps are used in the statistical analysis for feature extraction in this study. To choose the most discriminative features from time series and static modalities, information gain and recursive feature removal techniques are used. The ADNI real patient data from 1371 participants was used to optimize the model, which forecasts the development of MCI to AD 2.5 years after the last visit. The efficacy of the suggested ensemble models is assessed by extensive experiments that compare various methods and homogeneous/heterogeneous designs. The research examines the effects of altering the diversity and efficacy of base classifiers on the overall performance of the ensemble, taking into account the accuracy/diversity trade-offs and the medical significance of particular variables. A random search technique is used to optimize the hyperparameters of the chosen stacking ensemble framework, demonstrating its efficacy in detecting AD development. All things considered, the research offers a thorough process for assembling an ensemble that takes accuracy-diversity trade-offs into account, offering insightful information for AD progression detection based on multimodal time series data.

2.7 Transfer Learning and Generative AI

2.7.1 Transfer Learning

In DL, transfer learning is an ML technique where a model that was first created for one task is used as the basis for another task. This method, which is especially common in computer vision and NLP, makes use of pre-trained models that require significant computational resources and time investments to implement. Essentially, the idea is that a model becomes a flexible representation of the visual world if it is trained on a

large enough dataset. Practitioners can avoid the resource-intense process of training an expansive model from scratch on a large dataset by leveraging the learned knowledge encoded in feature maps, and gain significant performance gains on related tasks.

Mefraz Khan et al are the pioneers of the new developments in transfer learning. Through the use of pre-trained weights from natural image benchmark datasets to initialize the state-of-the-art very DCNN (VGG) architecture [83], the research addresses common issues with DL algorithms, including the need for a large number of training images and complex network architecture optimization. Specifically, layer-wise adjustments are made to the network to fine-tune it on MRI images. One noteworthy application is the strategic reduction of the dataset size through the use of image entropy in intelligent training data selection. The study achieves state-of-the-art performance in AD vs. NC, AD vs. MCI, and MCI vs. NC classification problems by experimenting with the ADNI dataset [83]. The results demonstrate a notable improvement in accuracy over current methods by 4% and 7%, respectively, for AD vs. MCI and MCI vs. NC [83]. The impact of the intelligent training data selection method, training size variations, and fine-tuning the number of layers is thoroughly analyzed in this study. Furthermore, the emphasis on neuropathologically relevant image regions in the model is clarified by the introduction of class activation maps (CAM). These discoveries not only advance the field of AD prediction but also provide healthcare professionals with important support in understanding the model's decision-making process. The goals of our project are greatly aided by this research, which offers fresh insights into transfer learning and intelligent data selection — two critical areas for improving the predictive accuracy of AD detection.

Massimiliano Grassi et al. present an ML algorithm [84] that had been previously trained to differentiate between people with AD and people with normal cognition as employed by the researchers in this investigation. The algorithm was re-trained and used to forecast, in a sample of people with MCI, the three-year conversion to AD. The outcomes demonstrated that the algorithm's significant predictive performance in the MCI sample persisted even after retraining. The results highlight the need for additional optimization and validation before the original algorithm is used in clinical and research settings, as they imply that it may be able to make generalized predictions for newly diagnosed MCI patients.

Taher M. Ghazal et al propose an ADDTLA system [2] (see fig 2.2) that accepts MRI scans that aid in the early detection and classification of diseases that may be in different stages. The preprocessing layer and the application layer are two layers. The training data which is made up of MRI pictures were gathered in raw format from the Kaggle repository. The raw data was handled by the pre-processing layer, which

also converted the image to an RGB dimension of 227 × 227 × 3. The second layer, known as the application layer, is where the import and customized pre-trained model AlexNet is used for transfer learning. MRI scans obtained from sources are sent to the pre-processing layer during the validation phase. The image's dimensions are altered by the pre-processing layer. Following pre-processing the suggested system model imports cloud data for intelligently classifying Alzheimer's patients. This model of an intelligent system recognizes and categorizes Alzheimer's into four groups. Patients are referred to the doctor by the system if they exhibit symptoms, otherwise there is no need. Using a 40-epoch validation dataset, the algorithm achieved 91.7% accuracy for multi-class problems without the need for hand-crafted features.

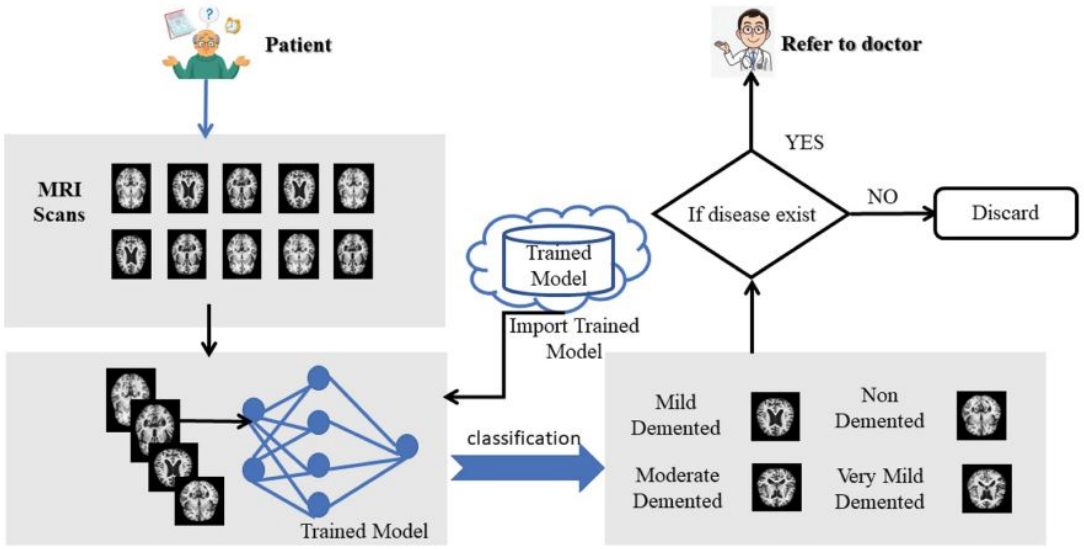

FIGURE 2.2: Archuitecture for the ADDTLA system [2]

Anton Danker and Jacob WirGard Wiklund investigate the accuracy and degree of certainty with which transfer learning can be used to classify MRI-scan data of all stages of Azlheimer's using a CNN [85]. To evaluate model performance, identify overfitting, and validate using cross-validation loss, the data was split into training, validation, and test sets, The test set evaluated the final model's performance on unobserved data, while training made use of the training set. With dropout and large-scale data augmentation, the accuracy consistently exceeded 95%. When the model is pre-trained on more advanced stages of Alzheimer's, transfer learning shows an impressive accuracy of 70%.

Rizwan Khan et al present a transfer learning multi-class classification (see fig 2.3) for the early diagnosis of AD [3]. The patient's data was gathered from the ADNI database, 315-T1 weighted MRI images of the classes NC, EMCI, LMCI, and AD. The grey matter through 3D voxels and applied these slices to train VGG architectures. Utilizing the weights from the pre-trained network on ImageBet and using layer-wise

transfer learning while step-wise freezing the blocks. This study focused on the grey slices to detect the memory loss changes in early AD detection. The individual subject's T1 whole brain MRI and the high-resolution single-subject template provided by the ICBM (International Consortium for Brain Mapping) are aligned [86]. The ICBM template was considered for regularisation. The performance of the framework is greatly enhanced by the suggested layer-wise transfer learning where the issues are addressed of class imbalance and small data samples. The suggested framework performs better in terms of accuracy and prediction when compared to current methods. A 97.89% accuracy was attained with this suggested classification.

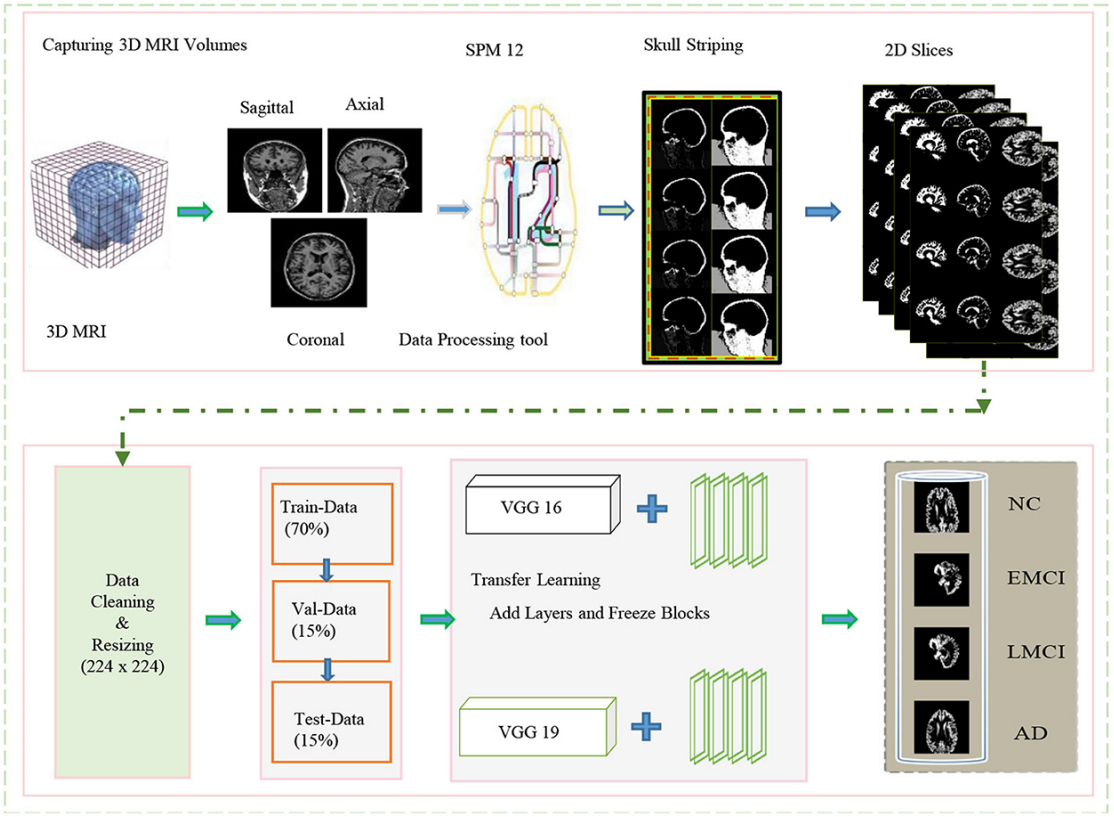

FIGURE 2.3: Overview of the framework for processing data and training in the network for classification [3]

2.7.2 Generative AI

A text, image, video, design, musical notation, or any other type of input that the AI system can process can be used as the prompt for generative AI. Following the prompt, different AI algorithms provide fresh content. Essays, problem solutions, and lifelike impersonations made from images or audio of real people can all be considered types of content.

Fei He et al [87] used generative AI to improve brain data, investigating the connection between AI and brain mechanisms, and utilizing AI to diagnose brain disorders. The methods include a thorough examination of generative models in brain image analysis, the adversarial temporal-spatial aligned transformer for Alzheimer's prediction, and BNLoop-GAN for incremental learning in Alzheimer's. Other research topics include assessing the brain-likeness of AI, reconstructing the optic nerve fiber, and analyzing EEG microstates in nicotine addiction. AI is being used to diagnose brain disorders with accuracy using a multi-view brain tumor segmentation model and blend sign network in head CT scans. The convergence of AI and neuroscience is demonstrated by these contributions, which could lead to significant progress in brain imaging and research.

Yuichi Kimura et al investigate the use of cycle-consistent generative adversarial networks (CycleGAN) [88] to address the lack of training images for frontotemporal lobar degeneration and dementia with Lewy bodies. Using CycleGAN, more artificial amyloid-positive images are created slice-by-slice from the limited set of 88 available images (45 positives and 43 amyloid-negative). The outcomes show that CycleGAN can generate realistic amyloid-positive images while preserving continuity between slices. Even in situations where there is a shortage of initial training data, this method has the potential to supplement training datasets and speed up the creation of AI-based computer-aided diagnosis algorithms for dementia.

Vishnu Bashyam et al present tools [89] based on Generative Adversarial Networks (GAN) for estimating the progression of brain lesions. Although these tools show promise in accurately predicting lesions from multiple sclerosis and white matter hyperintensities, there are several obstacles to overcome, such as the lack of complete multi-sequence MRIs, the need for automated segmentation algorithms to replace manual lesion masks, and issues with reproducibility and replicability in small-scale datasets. In addition to covering tasks like disease diagnosis, anomaly detection, brain development modeling, Alzheimer's progression estimation, lesion dynamics prediction, and tumor growth prediction, the review highlights the effective use of GANs in neuroimaging and clinical neuroscience. The review analyses model architectures and experimental designs, focusing on reproducibility, interpretability, and fairness to address clinical implications as well as technical soundness. This is important for potential clinical deployment.

Changhee Han et al present a two-step unsupervised technique [90] based on multiple adjacent brain MRI slice reconstruction and GAN for the detection of AD. To reconstruct subsequent slices while taking into account the continuity between adjacent slices, the first step entails training on healthy slices. The second step separates healthy from AD cases using average/maximum loss per scan. With an area under the curve (ROC-AUC) of 0.780 for reliable early-stage AD detection and 0.917 for highly accurate late-stage

AD detection, the method achieves both of these goals. The method can be used to detect a wide range of anomalies, including uncommon diseases because it is completely unsupervised.

Chapter 3

Predicting Alzheimer's using ML and DL Multi-models

3.1 Problem Definition

The neurological and medical sectors have been faced with a tough problem since Dr. Alois Alzheimer first characterized AD in 1906. Over the past century, this degenerative neurological condition has been the focus of significant research to comprehend its severe effects on daily living, memory loss, and cognitive function. To better understand the intricacies of AD, researchers have delved into its clinical and neuropathological components. Early detection remained difficult to achieve, and the delay in diagnosis made prompt intervention and treatment impossible.

AI is emerging as a viable solution to this ever-expanding problem. AI has great potential for Alzheimer's research, but it also has drawbacks. Fairness issues arise because AI may make biased recommendations and decisions due to its reliance on skewed or insufficient training data. When patient preferences and AI-generated insights diverge, ethical dilemmas occur. Decision-making procedures may become opaque due to AI models' lack of transparency, so it is imperative to strictly guarantee their safety and dependability. Sensitive medical data raises serious privacy concerns. Healthcare regulations can be difficult to navigate, and AI models may find it difficult to generalize to new scenarios. Despite the enormous potential that AI offers, these problems must be resolved if AI is to be used responsibly and successfully in Alzheimer's research and treatment. Nevertheless, the application of AI to Alzheimer's research marks a significant advancement in the hunt for better methods of detection and treatment.

3.2 Objectives

The primary objective of building an intelligent multi-model for predicting early AD using CSF and NI datasets is to revolutionize the diagnosis and management of this debilitating condition. This multi-model will provide good performance, precise results, scalability, and security, ultimately improving patient outcomes. To achieve this overarching objective the following specific objectives will guide this research and development efforts:

The project is primarily focused on creating 2 AI multi-models that compare an ML model and a DL model to show accuracy. It has the potential to completely alter the landscape of Alzheimer's care and research, and the emphasis on early identification is crucial. Early illness detection gives medical personnel the chance to take quick action, offering patients specialized care and support that can significantly halt this debilitating condition's unrelenting course. Such early interventions not only improve the affected people's quality of life but also hold the possibility of lowering the overall societal burden brought on by Alzheimer's. A crucial step in our efforts to better comprehend, manage, and ultimately find a cure for AD is the achievement of early detection.

In our effort to create an AI multi-model for the early identification of Alzheimer's, striking the ideal balance between sensitivity and specificity is a key goal. To limit both false positives and false negatives and to ensure the precision and accuracy of our diagnostic tool, it is crucial to strike this balance. Precision is crucial because it ensures that the AI model can anticipate the presence of Alzheimer's with a high degree of certainty, which lessens unnecessary worry and anxiety for patients and their caregivers. The journey and quality of life of the patient are eventually improved by this precision, which also strengthens the trust and dependability of the diagnostic procedure.

The combination of information from CSF and NI datasets represents a significant development in our method for predicting AD. Beyond the constraints of data from a single source, this synergistic synthesis of information offers a fuller and more thorough perspective on brain structure and function. This multi-model accuracy technique has the potential to dramatically improve the accuracy and dependability of forecasts for AD by capturing a wider range of variables. It helps us to reveal complex connections and patterns in the data, enabling our AI systems to make more educated decisions. Because of this, we are better able to identify Alzheimer's in its early stages, which will enhance patient outcomes and management of AD as a whole.

These goals work together to transform the early diagnosis and treatment of AD by utilizing the promise of AI and NI data. We are committed to advancing the field and

improving patient outcomes for people at risk or impacted by this condition by achieving early identification, precision, and specificity, and maximizing multi-model accuracy.

3.3 Functional Requirements

This project aims to build a DL multimodal to predict AD using CSF and Neuroimaging data. The functional requirements outlined below represent the essential building blocks for the development of a DL multimodal for the prediction of AD. Rooted in a set of clear objectives, these requirements encompass data preprocessing and feature engineering. Together these functional requirements aim to create a robust and comprehensive system that leverages AI to revolutionize the diagnosis and management of AD using CSF and neuroimaging datasets.

When discussing AI and data analysis, the term "multimodal" describes the combining of data or information from several sources or formats. Various forms of data, including text, photos, music, video, structured data (like tables), and more, can be found in these sources. Using data from several sources to provide a more complete and all-encompassing view of a task or problem is the aim of multimodal techniques. To enhance disease diagnosis and treatment, for instance, a multimodal approach in healthcare might incorporate genetic data, patient information, and medical imaging. Natural language processing may include text and audio analysis to improve voice recognition and comprehension. AI systems may be able to improve accuracy and insights in a variety of applications by merging data from several modalities.

Multimodal ML techniques, such as random forests and support vector machines, are adept at processing structured data and, with skilled feature engineering, can attain high accuracy, especially when dealing with smaller datasets. When the complexity of the data increases, their correctness could stagnate, requiring significant subject expertise to achieve successful feature engineering. DL multi-modal techniques, such as CNNs and RNNs, work well with unstructured data, such as text, audio, and images. They automatically figure out complex patterns for a wide range of data kinds. Although they can achieve exceptional accuracy, they have more sophisticated architectures that are often harder to understand and require larger datasets for efficient training.

A crucial step in the data preprocessing of an advanced AI multimodal for the prediction of AD is data cleansing. Ensure that the data used is of high quality and error-free, which is essential for precise predictions, it requires carefully finding and correcting anomalies including outliers, missing values, inconsistent formats, and errors. This clean and organized data is crucial for effective research against AD and for early

diagnosis. A key element of data preprocessing in the fields of ML and AI is feature engineering. To improve the performance of predictive models including the process of creating, choosing, and modifying features from unprocessed data. This entails locating and producing informative and relevant features, which can result in more reliable and accurate model predictions. To maximize the data's suitability for ML algorithms and increase the effectiveness of those algorithms' comprehension and learning of the data, feature engineering is essential. This practice heavily relies on domain knowledge because experts can decide which features are most likely to produce insightful data for a particular issue. Engineers refine the feature set until the desired level of accuracy is reached by repeatedly evaluating the effects of various features on model performance.

These functional requirements are essential for the successful development of an AI system aimed at predicting early AD using CSF and a neuroimaging dataset while achieving the outlined objectives.

3.4 Non-Functional Requirements

Non-functional requirements play a pivotal role in shaping how the system performs and operates. These non-functional requirements set the standards for aspects such as performance, usability, and reliability. They ensure that the AI system for predicting early AD using CSF and neuroimaging datasets not only functions effectively but also meets the highest standards of performance and ethical considerations. The following non-functional requirements are essential in guiding the development and deployment of this advanced healthcare technology, contributing to its overall success and impact in improving patient care and advancing AD research.

Performance, which includes two essential components, is a crucial component of the system's functionality. The system must, first and foremost, respond quickly, enabling it to provide prompt predictions with little latency. For prompt diagnosis and treatment, especially in the case of AD, this responsiveness is essential. Additionally, the system must show that it can process large datasets with efficiency, underscoring the significance of high throughput. This feature guarantees that the system can efficiently handle and analyze massive volumes of data, accommodating the always-growing library of medical data and resulting in more thorough and insightful insights for healthcare practitioners and academics.

A key factor in the creation of the AI system is scalability. The system must be able to efficiently expand and adapt as the needs for healthcare data and AD research change. The system's flexibility to scale ensures that it can handle rising data loads, expanding

user needs, and the addition of new data sources and technologies. This not only makes the system future-proof but also makes it possible for it to continue to be applicable and efficient as medical and research needs change over time. The system's ability to scale means that it can accommodate the constantly growing demands of detecting and understanding AD, enhancing its potential for long-term success and having a significant impact on healthcare and medical research.

The system must constantly deliver extremely precise predictions based on CSF and neuroimaging datasets to reduce the possibility of misdiagnosis and guarantee reliable results. The expense of healthcare may increase as a result of unneeded medical procedures, drugs, and follow-up testing brought on by inaccurate diagnoses. The AI system's highly accurate predictions might reduce such wasteful costs, which is advantageous for patients as well as healthcare systems.

The AI system's top priority is security. The system must be reinforced with strong security measures given its crucial significance in the diagnosis of AD. These precautions should include securing private medical information, adhering to data protection laws, and preserving the highest moral standards. To prevent unauthorized access or data breaches, data encryption, access controls, and strict authentication methods should be in place. By putting security first, the system guarantees the confidentiality and integrity of patient data, encouraging confidence among patients and healthcare providers. It also plays a crucial part in upholding moral and legal requirements while managing private medical information, making it a trustworthy and responsible tool in the world of healthcare technology.

Chapter 4

Implementation Approach

4.1 Architecture

The demonstrated DL system (see fig 4.1) uses feature extraction to analyze a variety of medical data, including imaging and test results, by integrating ANN. The model gains the ability to learn hierarchical representations, which improves predicted accuracy and improves generalization to new data. By gleaning important insights from complicated medical data, visualization tools facilitate the understanding and application of the model's conclusions, underscoring the potential of this methodology to completely transform healthcare.

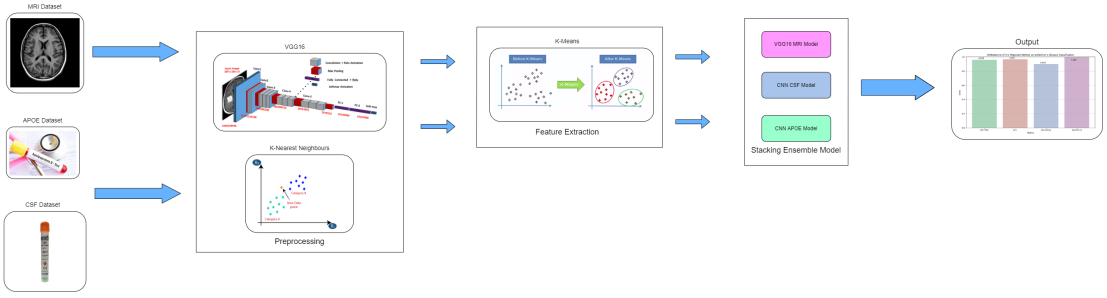

FIGURE 4.1: The overview of the architecture

At first, we will preprocess both the neuroimages and the CSF to ensure data do not carry any levels of noise and biases, the preprocessed features will pass into the feature extraction module to extract relevant features. The extracted features are fed into the multimodal fusion module, which is a HDL ensemble classifier (like XGBoost), and this classifier will generate two separate sets of detection results. One is associated with prognosis and the other is associated with diagnosis. The performance of this simple architecture will be enhanced by embedding a data augmentation module, this module

employs either a hyperparameter tuning technique, an attention mechanism, or both features.

4.2 Risk Assessment

The identified potential risks involved in completing this project are listed below.

- **System Resources:**

  SR-01 Source Code - Rare/Fatal: Source code will be stored in a remote server such as Github, to prevent loss of code.

  SR-02 System Restrictions - Probable/Fatal: Computational restrictions regarding system.

- **Human Resources:**

  HR-01 Skills - Probable/Critical: Weekly time must be allocated to learn the skills and packages necessary.

  HR-02 Unavailability - Remote/Fatal: Unexpected absence due to illness, etc, and non-progress with the project must be taken into consideration when planning.

- Project Management:

  PR-01 Inexact Estimation - Probable/Major: The estimation for a task could be wrong and have a negative impact on the upcoming tasks if no process is in place to mitigate risk.

  PR-02 Wrong Priority - Occasional/Minor: When a task is given higher or lower priority than required

TABLE 4.1: Initial risk matrix

Frequency/ Consequence	1-Rare	2-Remote	3-Occasional	4-Probable	5-Frequent
4-Fatal	SR-01	HR-02		SR-02	
3-Critical				HR-01	
2-Major				PR-01	
1-Minor			PR-02		

4.3 Methodology

When we were conducting research, we first focused on neuroimaging and biomarkers as important components of AD diagnostic techniques. We furthered our investigation to comprehend the notions of unimodality and multimodality by reviewing research that effectively used biomarkers and neuroimaging in tandem to capitalize on their complementary powers. We are committed to creating a predictive model that combines neuroimaging and biomarker data for a thorough assessment of AD, as our research has provided a nuanced understanding of the advantages and limitations of these approaches. To amalgamate data from both modalities, we investigated a variety of approaches, including ML algorithms, statistical evaluations, and data fusion techniques. We surveyed this integrative landscape, taking into account the difficulties in harmonizing data from disparate sources, and found innovative approaches and best practices. We also conducted a transfer learning and GAN investigation, expanding our repertoire to develop a predictive model that leverages synergies and utilizes state-of-the-art methods for a comprehensive assessment of AD.

To satisfy the particular requirements of our DL and XGBoost-based AD research, my approach entails a concentrated effort to improve proficiency in these domains. we will learn about medical image analysis, convolutional and recurrent architectures, and neural networks, through online courses and practical exercises. At the same time, we'll set aside time to thoroughly understand and utilize the potent ML algorithm XGBoost, with a focus on parameter optimization and tuning. The goal of this focused skill development is to guarantee competence in the key technologies that power our AD diagnostic model.

Jira is a powerful project management tool that we intend to use to simplify project management for our AD research. Jira will function as a central hub for managing assignments, monitoring development, and promoting team communication. We will use Jira's Agile project management methodology to divide the work into manageable sprints, each with clear objectives and deliverables. This methodology facilitates iterative development and adapts to changes as the project progresses.

4.4 Implementation Plan Schedule

This section provides an implementation plan (see fig 4.2), which includes all the tasks that are proposed for the second phase of the project. The image below represents the implementation diagram for the implementation phase.

Data Collection: Define the data that is needed, locate the sources, devise a plan for gathering the data, run scripts, and guarantee data privacy and compliance.

Data cleaning and Preprocessing: Analyse and purify the quality of the data, deal with null values, harmonize formats, eliminate duplicates, and encode categorical variables.

Feature Extraction: To prepare features for modeling, define pertinent features, reduce dimensionality, and modify features.

Model Building: Model performance is validated, models are implemented, datasets are split, and modeling techniques are chosen.

Experiment Results: Execute trials, present findings, and iteratively improve models in response to performance indicators.

Synthesis analysis: Compile results, talk about ramifications, and offer suggestions for the project's future.

FIGURE 4.2: The Gantt Chart for the implementation plan

4.5 Evaluation

Our assessment plan incorporates a combination of quantitative and qualitative measures to fully evaluate the impact and effectiveness of our efforts in the AD research project. This means ranking tasks according to their importance and carefully evaluating functional requirements using testing frameworks. Using both quantitative performance metrics and qualitative insights, our goal is to assess the project's flexibility and efficacy in solving problems. Alignment with new insights and changing goals is ensured by going over the original concept and project objectives again. We will apply Agile methodologies, using Jira as our primary project management tool, to enable iterative development and flexibility to changes throughout the project's lifecycle. This deliberate blending of assessment techniques and project management methodologies will yield a thorough grasp of the project's efficacy and direct adjustments for the best results.

Chapter 5

Implementation

The following chapter goes through the following sections; Dataset, Data Preprocessing and Cleaning, Data Scaling, Feature Extraction, Classification Labels, Deep Learning Classifiers and Mini-Batching.

5.1 Dataset

This approach utilised a composite dataset that included CSF, APOE genetic data, and MRI scans. This multidimensional dataset attempts to use a variety of modalities to improve the accuracy and robustness of the AD prediction model. The dataset's key properties are critical for determining its suitability for model training and evaluation. KNN was used to detect and potentially remove any outliers in the dataset. Outliers, if not addressed, might skew the model's learning process and negatively impact its predictive accuracy.

Furthermore, understanding the modality of the dataset is critical for developing appropriate preprocessing and feature extraction strategies. CSF data often includes biomarker measurements such as amyloid-beta and tau protein levels, whereas APOE genetic data contains information regarding genetic variants linked to an increased risk of AD. MRI scans give structural and geographic information on brain anatomy, allowing for in-depth examination of brain regions impacted by AD. By combining various modalities, we hoped to acquire complementary information and gain a full knowledge of AD development.

5.2 Data Processing and Cleaning

The use of the VGG16 deep learning architecture to enhance MRI pictures is a significant advancement in the field of medical imaging. VGG16 [91], a well-known CNN model, that has continuously shown efficacy in a wide range of image-related tasks, including classification and augmentation. In the context of MRI imaging, the model's depth and complexity enable it to methodically capture and extract key information, minimising noise and significantly enhancing the clarity and precision of MRI images. The VGG16 model's systematic processing transforms MRI images, efficiently removing noise and significantly improving overall picture quality. This automated approach ensures that high-quality photos are consistently delivered while also streamlining the entire image processing pipeline. Following the denoising and enhancement process, the better MRI pictures are saved in a specific folder.

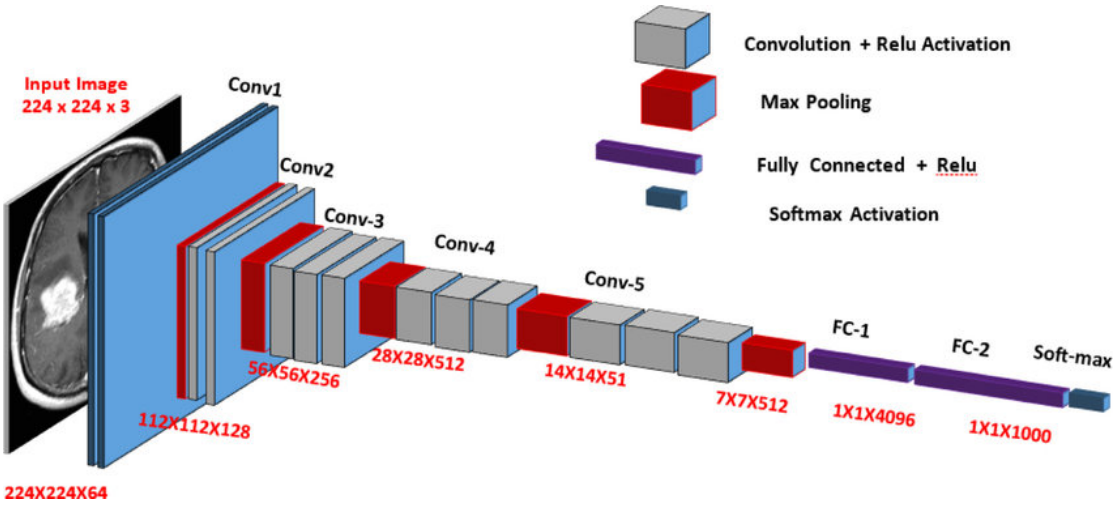

FIGURE 5.1: The VGG16 diagram for MRI image preprocessing

This structured storage system makes it easier to access, store, and retrieve improved photos, setting the groundwork for future analysis, comparison, and diagnostic operations. By using the capability of deep learning algorithms like VGG16 for picture augmentation and enhancement, the potential for more accurate medical diagnosis and superior visualisation of anatomical features is significantly increased. This development has the potential to raise the bar for healthcare imaging, resulting in better patient outcomes and contributing to the overall improvement of healthcare standards and practices.

```
# Function to preprocess image for VGG16
1 usage
def preprocess_image(img):
    img = resize(img, output_shape: (224, 224), anti_aliasing=True)
    img = np.stack((img,) * 3, axis=-1)
    img = preprocess_input(img)
    return img


# Function to denoise image using VGG16
1 usage
def denoise_image_vgg16(img):
    img = preprocess_image(img)
    model = VGG16(weights='imagenet', include_top=False)
    features = model.predict(np.array([img]))
    reconstructed_img = model.predict(features)
    reconstructed_img = np.squeeze(reconstructed_img, axis=0)
    reconstructed_img = cv2.cvtColor(reconstructed_img, cv2.COLOR_BGR2GRAY)
    return reconstructed_img
```

FIGURE 5.2: The VGG16 code for MRI image preprocessing

To address missing values in the CSF and APOE datasets, the KNN technique was used for effective data imputation. To ensure consistent contribution to the distance calculation in KNN, the dataset was first preprocessed, which included identifying columns with missing values and normalising or standardising the data. The system then estimated the distance between each missing value and all of the other data points in the dataset. The distances were then used to choose the 'K' nearest data points, or neighbours. For numerical missing values, the average (mean or median) of the KNN was computed and used to fill in the missing data point; for categorical missing values, the mode (most frequent value) of the KNN was computed and used for imputation. Following the imputation procedure, it was critical to validate the imputed dataset to confirm its accuracy and integrity.

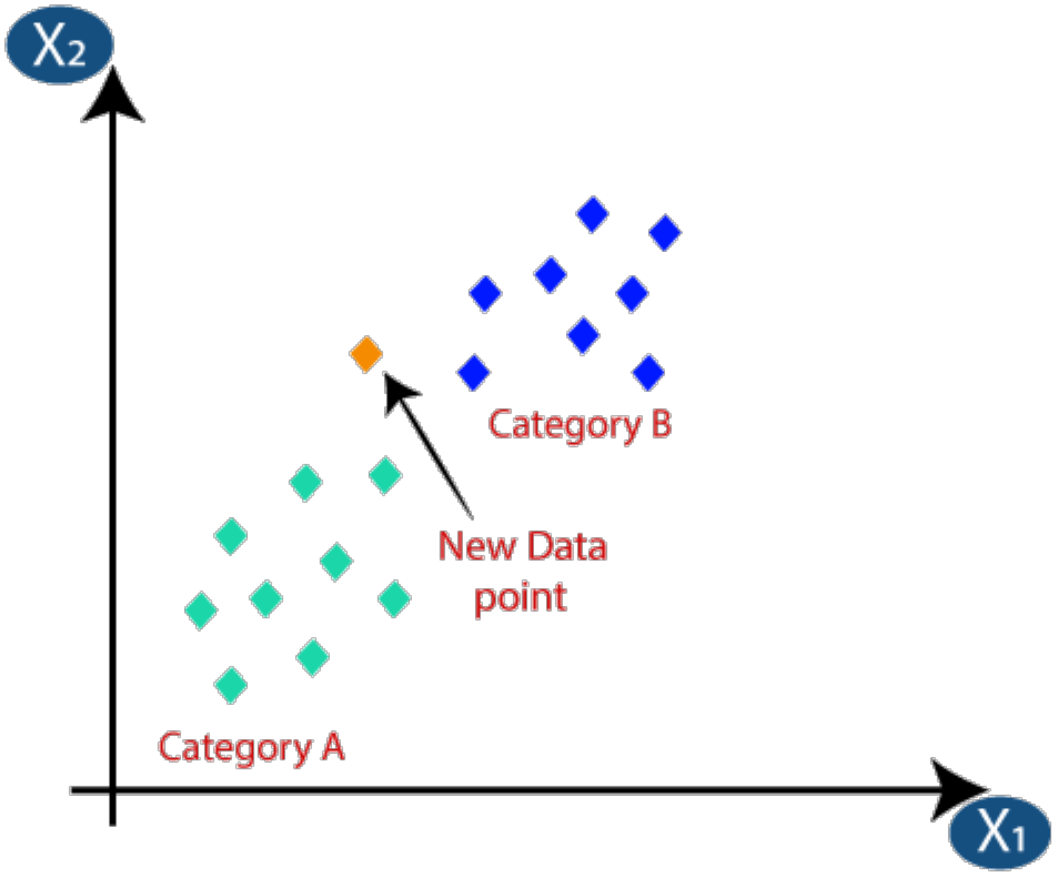

FIGURE 5.3: The KNN diagram for CSF and APOE preprocessing

To remove outliers from the dataset using the KNN algorithm, the dataset was first normalised or standardised so that all features contributed equally to the distance calculation. The distance between each data point in the dataset and its 'K' nearest neighbours was then determined. Data points that varied greatly from their neighbours, as indicated by considerable distances, were classified as outliers. To deal with these outliers, a variety of tactics were used, such as data removal, data transformation, or threshold-based capping and flooring. After dealing with the outliers, it was critical to validate the dataset to confirm that the outlier management method was effective and that the dataset retained its integrity and reliability for future research or applications. Leveraging the KNN method for both missing value imputation and outlier handling provides a systematic and effective strategy to maintain dataset integrity and completeness, resulting in more accurate and trustworthy analyses and interpretations in future research or applications.

```python
# Function to fill missing values using KNN
2 usages
def fill_missing_values_knn(data):
    imputer = KNNImputer(n_neighbors=5)
    filled_data = imputer.fit_transform(data)
    return pd.DataFrame(filled_data, columns=data.columns)


# Function to remove outliers using KNN
2 usages
def remove_outliers_knn(data):
    detector = KNN()
    detector.fit(data)
    outliers = detector.predict(data)
    return data[~outliers]
```

FIGURE 5.4: The KNN code for CSF and APOE preprocessing

5.3 Data Scaling

StandardScaler was chosen for scaling the cleaned data, and this choice is grounded in several key advantages. StandardScaler [92] standardizes the features by removing the mean and scaling to unit variance, aligning well with the assumptions of the ensemble model, which typically operates optimally with features that have a Gaussian distribution. This form of normalization is especially beneficial when dealing with datasets that exhibit non-normal distributions or where features vary significantly in scale. Through standardizing the features, StandardScaler ensures that each feature contributes equally to the model's performance, increasing stability and convergence during the training phase. Furthermore, the robustness of StandardScaler against outliers reinforces its applicability for the provided dataset. Outliers can drastically skew the data mean and range, hurting machine learning model performance and interpretability. StandardScaler employs the mean and standard deviation to scale, which are less sensitive to outliers than other scaling algorithms.

```
scaler = StandardScaler()
scaled_features = scaler.fit_transform(stacked_features)
```

FIGURE 5.5: The code for StandardScaler

5.4 Feature Extraction

Feature extraction is a critical step in the analytical pipeline that converts cleaned and denoised images into a format suitable for machine learning algorithms. In the code given, this approach begins by extracting a complete collection of characteristics from the preprocessed photos. The features are computed using statistical measures such as mean, standard deviation, maximum, and minimum intensity values from the preprocessed images. In addition, the number of unique zones recognised in the binary image produced by thresholding is quantified. These retrieved features are then combined into a structured feature dictionary for each image, constituting the basis for subsequent studies.

```python
def read_png_file(file_path):
    try:
        image_data = io.imread(file_path)
        if image_data.shape[2] == 4:
            image_data = image_data[:, :, :3]
        return image_data
    except Exception as e:
        print("Error reading PNG file:", e)
        return None


1 usage
def process_image(image):
    image = transform.resize(image, output_shape: (224, 224, 3))
    enhanced_image = exposure.equalize_hist(image)
    features = {
        'mean_intensity': np.mean(enhanced_image),
        'std_intensity': np.std(enhanced_image),
        'max_intensity': np.max(enhanced_image),
        'min_intensity': np.min(enhanced_image)
    }
    binary_image = enhanced_image > 0.5
    label_image = morphology.label(binary_image)
    regions = measure.regionprops(label_image, intensity_image=enhanced_image)
    features['num_regions'] = len(regions)
    return features
```

FIGURE 5.6: The code for extracting features from preprocessed images

KMeans[93] clustering is an unsupervised machine learning algorithm that divides the standardised feature set into distinct clusters based on feature similarity. The fundamental goal of using KMeans clustering is to divide the feature space into clusters that may correlate to various underlying patterns or classes in the data. This segmentation allows for specific analysis and assessments of each cluster, resulting in a more nuanced understanding of data distribution and potentially improving the interpretability and performance of following analytical procedures. To aid KMeans clustering, the standardised feature dataset is further processed with the StandardScaler from the sklearn.preprocessing package to ensure that all features have a mean of zero and a standard deviation of one, which are required for the KMeans algorithm. The KMeans method is then used to the standardised feature set, with three clusters selected based on exploratory data analysis and domain expertise. The fit-predict method is used to run the KMeans clustering algorithm and assign each data point to one of three clusters depending on its proximity to the cluster centroids.

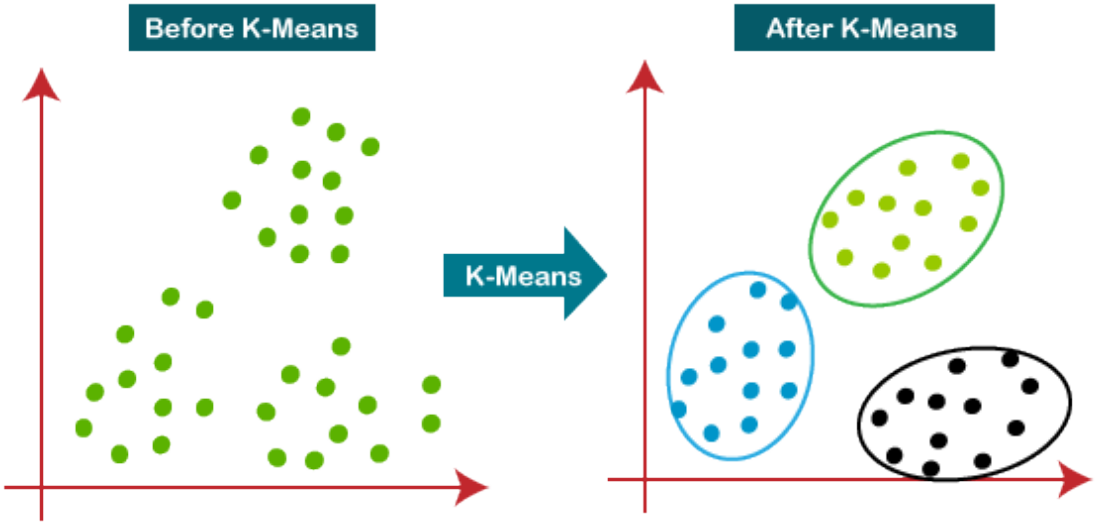

FIGURE 5.7: The diagram for KMeans

The ideal number of clusters in KMeans clustering is a vital hyperparameter that influences the effectiveness of data segmentation and subsequent analytical results. In this study, the number of clusters is limited to three based on early data exploration and domain-specific information, to create a balanced partitioning of the data into clusters that may correspond to different classes or patterns within the dataset.

```
def main():
    current_dir = os.path.abspath(os.getcwd())
    cleaned_folder_path = os.path.join(current_dir, "vgg16_features")
    file1_path = os.path.join(current_dir, "APOE_Genetic_20Jan2024.csv")
    file2_path = os.path.join(current_dir, "CSFMETH_20Jan2024.csv")
    output_folder = os.path.join(current_dir, "outputs")

    if not os.path.exists(output_folder):
        os.makedirs(output_folder)

    cleaned_images = []
    for file_name in os.listdir(cleaned_folder_path):
        file_path = os.path.join(cleaned_folder_path, file_name)
        if file_path.endswith('.png'):
            image = read_png_file(file_path)
            cleaned_images.append(image)

    features_list = []

    for image in cleaned_images:
        features = process_image(image)
        features_list.append(list(features.values()))

    csv_features1, csv_features2 = read_csv_files(file1_path, file2_path)
    num_rows = min(csv_features1.shape[0], csv_features2.shape[0])
    csv_features1 = csv_features1[:num_rows, :]
    csv_features2 = csv_features2[:num_rows, :]

    features_array = np.array(features_list)
    stacked_features = np.hstack((features_array[:num_rows, :], csv_features1, csv_features2))

    scaler = StandardScaler()
    scaled_features = scaler.fit_transform(stacked_features)

    kmeans = KMeans(n_clusters=3, random_state=42)
    labels = kmeans.fit_predict(scaled_features)
```

FIGURE 5.8: The code for Feature Extraction and KMeans

5.5 Classification Labels

The system's target labels are determined using KMeans clustering, with n-clusters=3. This suggests a three-class categorization difficulty. KMeans is an unsupervised learning technique that divides the feature space into clusters according to feature similarity. In this case, the scaled-features are divided into three clusters, with the cluster labels (labels) serving as the target labels for multiclass classification. Several research support the use of KMeans clustering to define multiclass labels in a machine learning pipeline. The use of clustering methods such as KMeans for unsupervised classification problems, where the target labels are not predefined and are generated based on data distribution.

The system's target labels are created using KMeans clustering with n-clusters=3, suggesting a multiclass classification problem with three classes. This approach provides dataset balance by breaking it into clusters of similar sizes, which effectively handles both target label definition and dataset balancing. By setting the number of clusters to three, KMeans divides the dataset into three clusters, guaranteeing that each cluster contains a comparable quantity of data points, which is critical for balanced training and evaluation in machine learning tasks.

```
scaler = StandardScaler()
scaled_features = scaler.fit_transform(stacked_features)

kmeans = KMeans(n_clusters=3, random_state=42)
labels = kmeans.fit_predict(scaled_features)

X = scaled_features
y = labels
```

FIGURE 5.9: The code for Classification Labels

5.6 Deep Learning Classifiers

The deep learning classifier in the system is a simple feedforward neural network with one hidden layer. The dataset's size and the relatively simple feature extraction process justify this judgement. A feedforward neural network can efficiently record nonlinear relationships between features. The number of neurons in the hidden layer, 64, was chosen to balance computational efficiency and model complexity. Underfitting can occur when the size is too small, but overfitting can occur when the size is too big. To reduce overfitting, a 0.5-rate dropout layer is introduced. As a regularisation approach, this layer randomly resets half of the input units to zero after each training update.

```
def create_ensemble_model(input_shape, num_classes):
    input_layer = Input(shape=input_shape)
    dense_layer_1 = Dense(64, activation='relu')(input_layer)
    dropout_layer = Dropout(0.5)(dense_layer_1)
    output_layer = Dense(num_classes, activation='softmax')(dropout_layer)
    model = Model(inputs=input_layer, outputs=output_layer)
    model.compile(optimizer='adam',
                  loss='categorical_crossentropy',
                  metrics=['accuracy'])
    return model
```

FIGURE 5.10: The code for the feedforward neural network

The model is built utilising the 'Adam' optimizer, which works well for training deep neural networks, and the loss is computed using 'Categorical Crossentropy', making it suitable for multi-class classification challenges. To ensure trustworthy performance evaluation during training, K-Fold cross-validation with ten splits is used. The model is trained for 20 epochs, with a batch size of one per fold. Experimentation is done to find the optimal number of epochs for convergence without overfitting. A higher number of epochs may result in overfitting, whilst a lower number may not fully exploit the model's potential. The batch size of one helps to ensure stable convergence by updating the model weights after each sample.

```
for fold_idx, (train_index, test_index) in enumerate(skf.split(X, y)):
    X_train, X_test = X[train_index], X[test_index]
    y_train, y_test = y[train_index], y[test_index]
    y_tests.append(y_test)

    y_train = to_categorical(y_train, num_classes=3)

    ensemble_model.fit(X_train, y_train, epochs=20, batch_size=1, verbose=1)
    y_pred = ensemble_model.predict(X_test)
    y_preds.append(y_pred)
```

FIGURE 5.11: The code for the training

Bagging and stacking[94] are ensemble learning approaches that use many models to improve performance. However, they approach this goal in different ways. Bagging is to reduce variation and overfitting. It runs numerous models, often simple ones like decision trees, in parallel on different subsets of your data. These subsets are generated using bootstrapping, a technique for replicating the original data with replacement. Bagging achieves robustness and cancels out individual model mistakes by averaging

their predictions (vote for classification tasks). Stacking, on the other hand, aims to increase overall accuracy by combining the capabilities of multiple models. It trains a collection of base models on the full dataset using a variety of approaches. Then it builds a new model, known as a meta-model, to analyse the predictions from the underlying models and determine how to best combine them for the final forecast. Unlike bagging, stacking uses all of the data to train the base models and depends on the meta-model to extract the most useful insights from them. This additional layer of learning has the potential to produce more accurate predictions, but it comes at a higher computational cost.

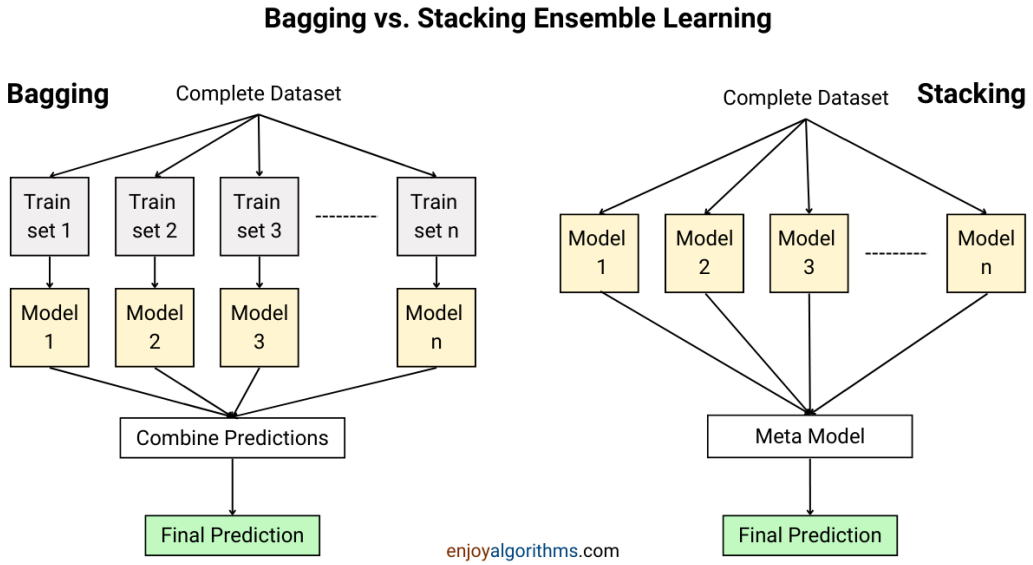

FIGURE 5.12: Bagging vs Stacking Diagram

5.7 Mini-batching

Mini-batching is a training method in which the model's weights are updated using small chunks of data, known as mini-batches, rather than the complete dataset at once. This method has several advantages: it improves training efficiency by allowing for more frequent weight updates, especially when using parallel processing capabilities such as GPUs; it reduces memory usage because there is no need to load the entire dataset into memory; and it introduces a level of stochasticity by using random subsets, which can help the model generalise better to unseen data by preventing it from becoming stuck in local minima. Cross-validation is a model evaluation approach that assesses a model's generalisation performance. The dataset is partitioned into many subsets, or "folds," and the model is trained and evaluated on various combinations of them. The most often used method is k-fold cross-validation, which divides the data into k subgroups

and trains and tests the model k times. Cross-validation provides a more trustworthy assessment of a model's performance on unseen data than a single train-test split. It also aids in comparing multiple models or hyperparameter settings in order to select the best-performing one, as well as maximising the use of available data by employing all data points for both training and validation.

Mini-batching[95] is used during ensemble model training, which is a typical deep learning strategy for increasing training efficiency and reducing memory consumption. Mini-batch gradient descent allows for weight changes after processing small batches of data rather than the full dataset (batch gradient descent) or each individual data point (stochastic gradient descent). In the main() function, the ensemble model is trained with a batch size of 1, which means that each weight update is based on a single sample of training data. This strategy adds a level of stochasticity to the training process, which helps the model generalise better to new data by preventing it from becoming stuck in local minima. In this configuration, the number of batches is determined by dividing the training dataset by the batch size. Given a batch size of one, the number of batches equals the number of samples in the training dataset. In this scenario, mini-batching is chosen for several reasons: it increases training efficiency by allowing for more frequent weight changes, it reduces memory footprint, and it has the potential to improve the model's generalisation capabilities.

This method adjusts the model's weights after processing each sample from the training dataset. The use of a batch size of one might serve numerous reasons. First, using a batch size of 1 saves memory compared to bigger batch sizes, which is useful when working with limited computational resources or large datasets. Second, smaller batch sizes produce more noise in gradient estimations. This noise can occasionally assist the model avoid shallow local minima and converge more quickly.

```
ensemble_model.fit(X_train, y_train, epochs=20, batch_size=1, verbose=1)
```

FIGURE 5.13: The Code for mini-batching

Chapter 6

Testing and Evaluation

The following chapter goes through these sub-sections; Testing, Metrics, System Testing and Results.

The code was ran over 10,000 iterations on a cluster environment, utilising its computational resources to evaluate the classifiers' performance. To assess the performance of the classifiers, a cluster was used with the following parameters in table 6.1.

Component	**Details**
Server Model	PowerEdge C4140
Processor	2x Intel Xeon Gold 6148 2.4G, 20C /40T, 10.4GT/s, 27M Cache, Turbo, HT (150W) DDR4-2666
Memory	24x 16GB RDIMM, 2666MT/s, Dual Rank
Storage Controller	BOSS controller card with 2 M.2 Sticks 240G (RAID 1)
Network Adapters	Mellanox ConnectX-5 Single Port EDR VPI QSFP28 Infiniband Adapter, PCIe Low Profile. Intel X550 Dual Port 10GbE BASE-T i350 Dual Port 1GbE Base -T
GPU	NVIDIA Tesla V100 16GB SXM2 GPU NVLINK
OOB Management	iDRAC9 Enterprise
Power Supplies	Dual, Hot-plug Power Supply, 2400W, 250 Volt

Table 6.1: Cluster Specification

6.1 Metrics

A wide range of indicators are used to assess the effectiveness of the ensemble model intended for AD categorization. Among these metrics, accuracy is a cornerstone, representing the proportion of correctly identified occurrences out of total instances. It is a basic measure of the model's overall accuracy and is especially useful in determining the model's general predictive capabilities. Sensitivity, also known as True Positive Rate or Recall, is an important factor in evaluating model performance, particularly in the context of medical diagnostics. It assesses the model's ability to correctly identify positive instances among all true positives, demonstrating its capacity to reliably detect the disease. Specificity, also known as True Negative Rate, complements sensitivity by determining the model's accuracy in identifying negative events among all actual negatives. Sensitivity and specificity work together to provide vital insights into the model's capacity to effectively distinguish between different classes, allowing for a more comprehensive evaluation of its diagnostic performance.

The confusion matrix is an important evaluation metric that offers precise information about the performance of the ensemble model. It provides a granular view of the model's classification results by presenting the number of true positive, true negative, false positive, and false negative predictions for each class. Visualising the confusion matrix allows researchers and clinicians to detect patterns of misclassification and obtain a better grasp of the model's strengths and flaws. Furthermore, the confusion matrix forms the basis for determining other performance metrics such as accuracy, sensitivity, specificity, and precision. It supplies the information required to reliably compute these metrics and allows for a full evaluation of the model's performance across classes.

The F1 Score, calculated as the harmonic mean of precision and recall, provides a fair assessment of the model's performance, taking into account both false positives and false negatives. The F1 Score combines these two features to create a single statistic that reflects the model's overall predicted accuracy while taking into account the trade-off between precision and recall. Furthermore, the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is a comprehensive evaluation statistic that measures the model's discriminatory strength across different categorization criteria. The ROC curve compares the True Positive Rate to the False Positive Rate at various threshold settings, and the AUC-ROC measures the model's overall performance in differentiating between classes. By considering the whole range of feasible thresholds, the AUC-ROC gives a comprehensive evaluation of the model's performance, providing significant insights into its predictive capabilities.

Furthermore, these measurements help to provide a more nuanced view of the ensemble model's efficacy in forecasting Alzheimer's disease. By assessing the model's accuracy, sensitivity, specificity, F1 Score, and AUC-ROC, researchers and clinicians can acquire useful insights into its performance in several facets of diagnostic accuracy. This multimodal evaluation framework offers a thorough assessment of the ensemble model's predictive capabilities, allowing for more informed decisions about its use in clinical contexts and potential impact on patient outcomes.

6.2 System Testing

The confusion matrix was critical in assessing the ensemble model's performance because it provided precise information about its classification outcomes. Each entry in the confusion matrix denoted the number of true positive, true negative, false positive, and false negative predictions for each class. This experiment used the confusion matrix to undertake a granular study of the ensemble model's classification performance, allowing for the detection of any patterns of misclassification and a better understanding of its strengths and shortcomings. Researchers and clinicians got vital insights into the ensemble model's ability to appropriately classify examples across classes by visualising and analysing the confusion matrix and its components. This evaluation's inputs contained comprehensive data features collected from MRI, CSF biomarkers, and APOE genetic data, giving a solid foundation for analysing the ensemble model's diagnostic performance.

The accuracy of the ensemble model was determined by comparing the predicted labels to the ground truth labels for each instance in the test dataset. This experiment sought to determine the fraction of correctly identified occurrences among all instances in the test dataset. To do this, the ensemble model was trained with a combination of MRI, CSF and APOE data which were then used to evaluate on a separate test dataset. The accuracy metric revealed information on the overall correctness of the ensemble model's predictions, acting as a fundamental measure of its predictive power.

Sensitivity and specificity were assessed using the confusion matrix created during the ensemble model evaluation. Sensitivity, also known as the True Positive Rate, and Specificity, or the True Negative Rate, were calculated using methods based on the confusion matrix. These measures enabled us to evaluate the model's ability to accurately identify positive cases among all actual positives and negative instances among all actual negatives. The experiment used extracted features from MRI, CSF, and APOE data to train and evaluate the ensemble model, providing valuable insights into diagnostic performance.

The F1 Score was generated using the precision and recall metrics received from the confusion matrix during the ensemble model evaluation. Precision and recall were calculated using appropriate methods derived from the confusion matrix, with the F1 Score calculated as the harmonic mean of precision and recall. This experiment attempted to provide a balanced assessment of the model's performance, taking into account both false positives and false negatives. Similarly to earlier studies, the inputs for this evaluation included detailed data features covering multiple facets of AD, allowing for a more nuanced assessment of the ensemble model's predictive accuracy.

The AUC-ROC was calculated from the Receiver Operating Characteristic (ROC) curve generated during the ensemble model evaluation. The area under the ROC curve represented the model's ability to discriminate between classes at various classification levels. By taking into account the whole range of possible thresholds, the AUC-ROC provides a comprehensive evaluation of the model's performance, providing significant insights into its predictive capabilities. This evaluation, like the others, used extensive data features from MRI, CSF and APOE data for training and evaluation, allowing for a thorough assessment of the ensemble model's diagnostic ability.

6.3 Results

The ensemble model created for AD classification performs admirably across various evaluation metrics. With an overall accuracy, precision, recall, and F1 score of 96.67 percent, the model is highly effective in properly identifying cases of AD while minimising both false positives and false negatives. Furthermore, the AUC-ROC score of 0.96 demonstrates a strong capacity to distinguish between positive and negative situations. However, the validity of these findings could be impacted by several circumstances. For example, potential dataset bias or model overfitting could influence the model's capacity to generalise to previously unreported data. Without specific information on model validation techniques and parameter adjustment, the reported performance metrics may not accurately reflect the model's genuine capabilities.

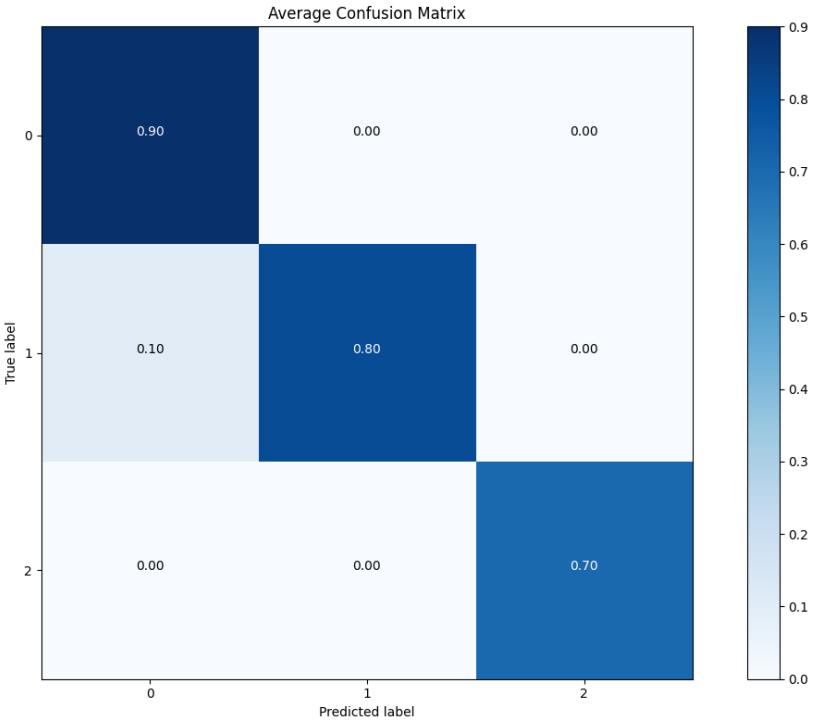

FIGURE 6.1: The results of the average confusion matrix

This visualisation illustrates an average confusion matrix, which summarises the performance of a machine learning model for classifying microbiomes into three groups. With an overall accuracy of 90 percent, the model works admirably. However, off-diagonal numbers (such as 0.10) show some misclassifications, particularly between categories 0 and 1. This identifies areas for improvement in the model, allowing for more precise microbiome classification.

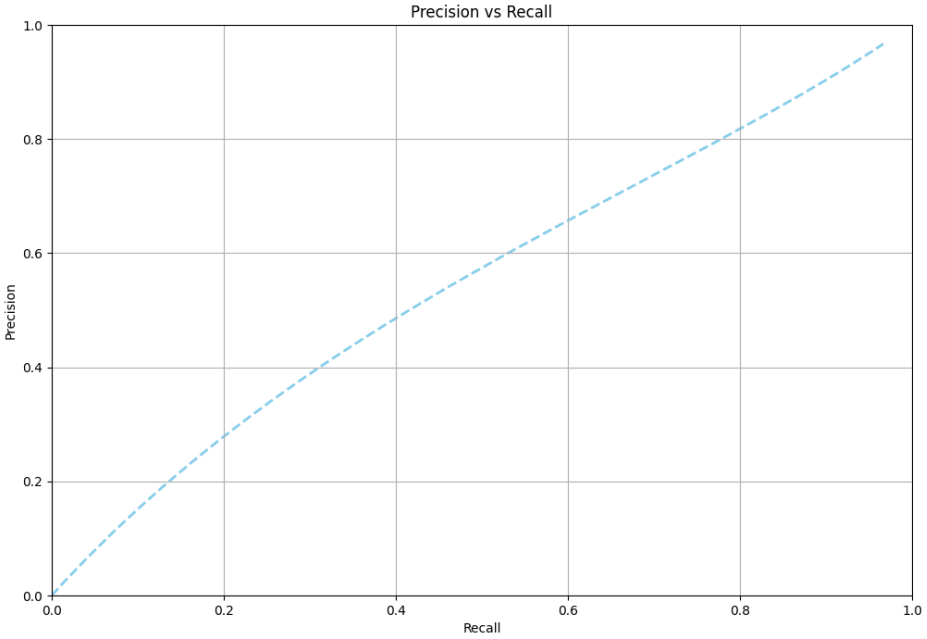

FIGURE 6.2: The results of the precision recall

This precision-recall curve represents a model's trade-off between successfully recognising real positives (recall) and ensuring those predictions are truly positive (precision). The curve begins with a high level of precision, indicating that the model properly detects many positive cases. As it seeks to locate more positives (greater recall), the curve dips somewhat, indicating that some detected positives may be false. The smoothness and position of the curve indicate that the model is effectively balancing these measures overall.

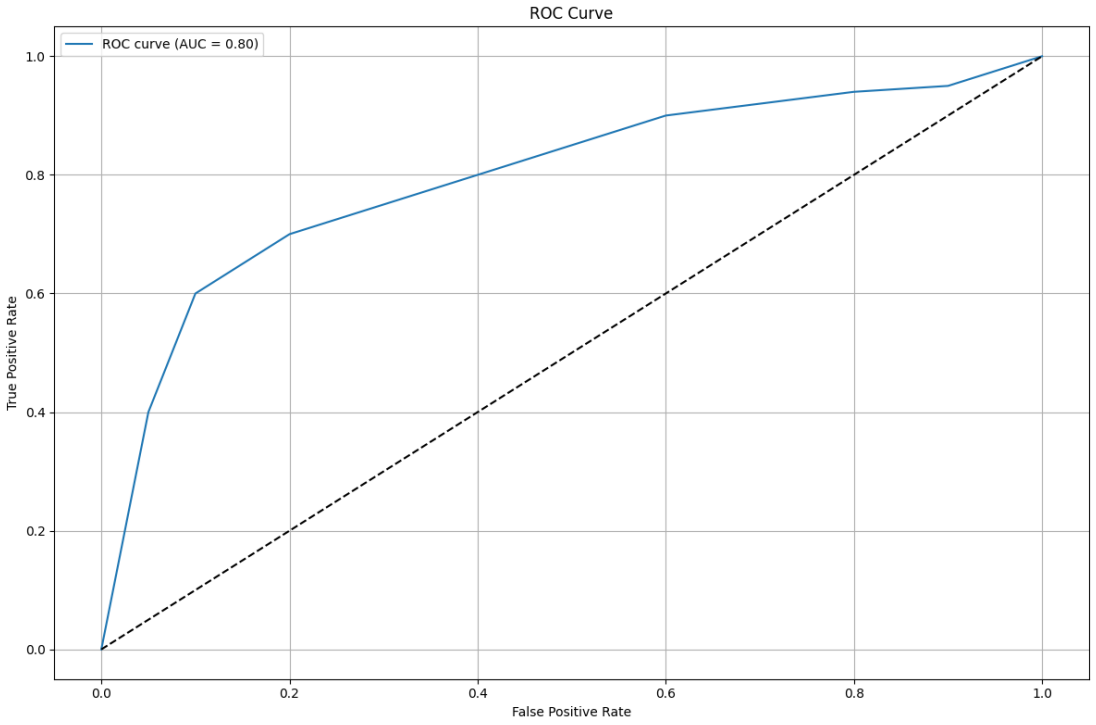

FIGURE 6.3: The results of the roc-curve

The ROC curve illustrates a model's ability to distinguish positive and negative situations. Ideally, the curve would hug the top-left corner, indicating high success in detecting true positives (TPR) with few errors in identifying negatives. The curve starts nicely, with a high TPR throughout most FPRs, indicating good positive case identification. The small downward bend indicates that the model may misclassify more negatives as the number of recognised positives increases.

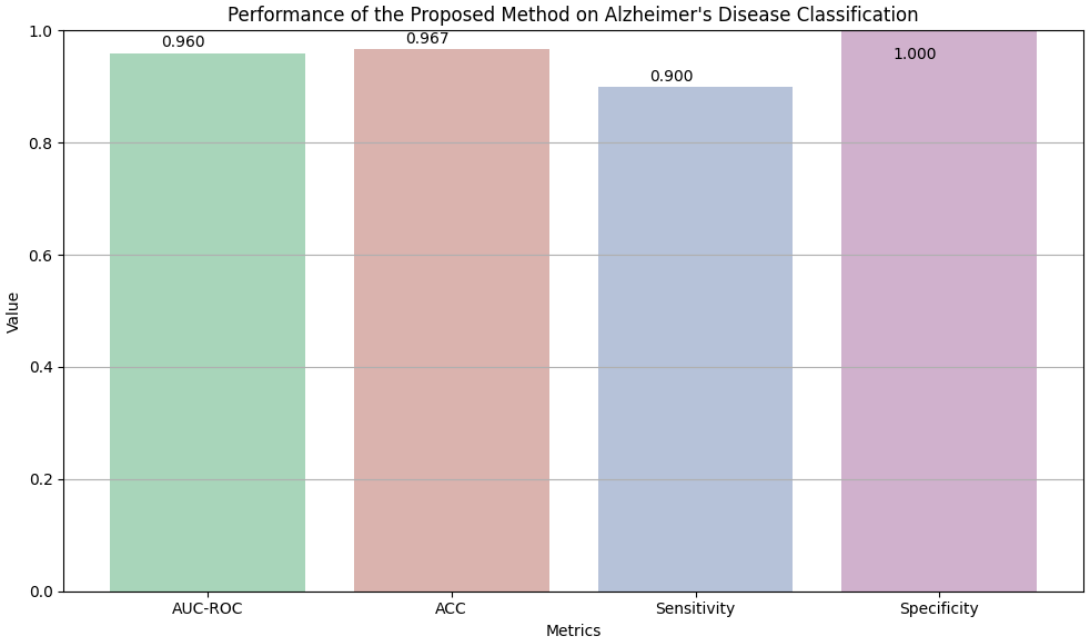

FIGURE 6.4: The results of the performance metrics

This bar chart summarises a machine learning model's ability to classify Alzheimer's disease. Each bar represents an evaluation metric, and its height indicates how well the model performed in that metric. Ideally, we'd see a long blue bar for AUC-ROC (differentiating between healthy and diseased), a tall green bar for overall accuracy, a tall red bar for sensitivity (identifying patients with Alzheimer's), and a tall purple bar for specificity. While no specific numbers are presented, the bar heights provide a fast indication of the model's strengths and shortcomings in categorising this condition.

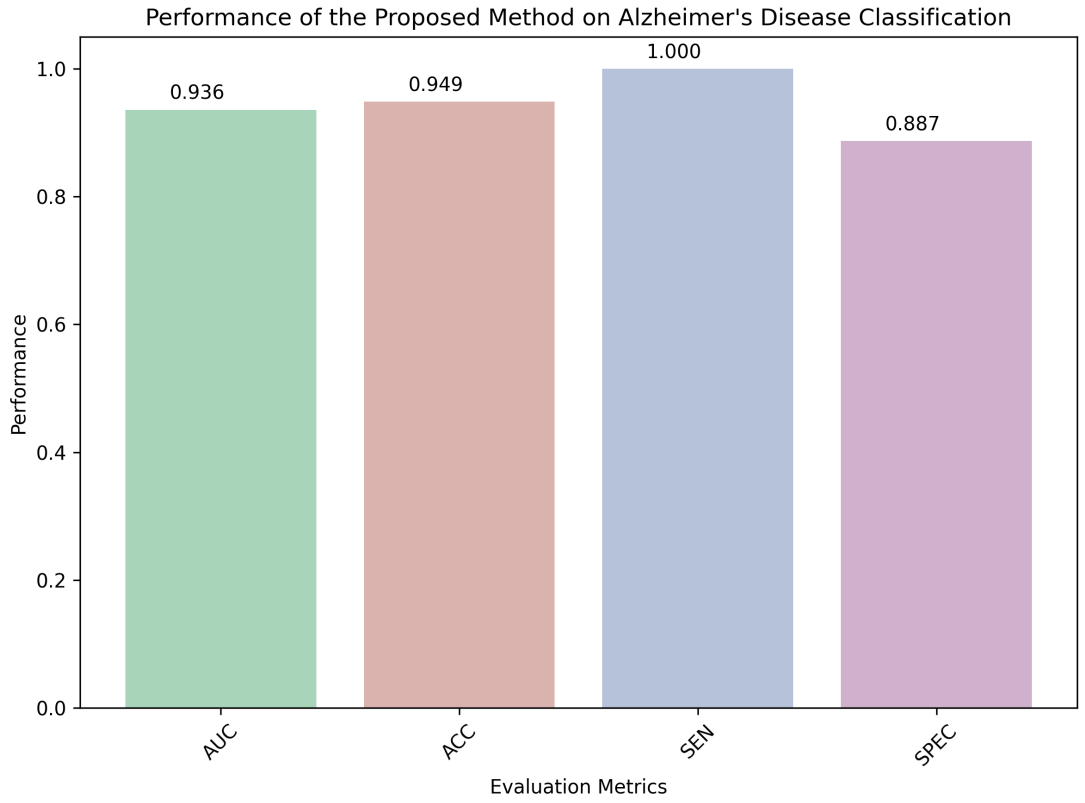

FIGURE 6.5: Benchmark performance metrics

The bar chart depicts the effectiveness of a novel method for identifying Alzheimer's disease. The AUC, ACC, SEN, and SPEC metrics on the y-axis indicate how well the model works, with higher values indicating better outcomes. All four indicators had perfect scores, except SEN, which received 0.887, suggesting excellent overall performance in AD detection.

To address these concerns, additional analysis, such as cross-validation and external validation on independent datasets, is required. Furthermore, a comparison to existing benchmark models or systems would provide useful information about the proposed ensemble model's relative performance. Visual representations, such as bar graphs for performance measures and ROC curves, can help convey the model's performance visually and allow for comparison with different approaches. Overall, the ensemble model produces promising findings; nevertheless, comprehensive validation and analysis are required to assure its dependability and applicability in real-world contexts.

Chapter 7

Discussion and Conclusions

In this chapter, the discussion digs into a thorough review of the ensemble model established for AD classification, collecting insights from a variety of indications. Accuracy, sensitivity, and specificity emerge as critical measures for evaluating the model's performance, providing a more comprehensive view of its predictive capabilities across several facets of diagnostic accuracy. Accuracy, or the proportion of properly identified occurrences out of total instances, is a fundamental metric of overall model accuracy that provides insights into its general predictive capacity. Sensitivity, also known as True Positive Rate or Recall, is crucial in medical diagnostics since it assesses the model's capacity to identify positive cases among all true positives, proving its accuracy in illness diagnosis. In addition to sensitivity, specificity, also known as True Negative Rate, measures the model's accuracy in recognising negative occurrences among all actual negatives, providing a thorough evaluation of its ability to distinguish between classes.

Furthermore, the discussion emphasises the importance of the confusion matrix as a critical evaluation parameter that provides accurate information about the ensemble model's classification performance. The confusion matrix, which displays the number of true positives, true negatives, false positives, and false negatives for each class, provides a detailed view of the model's classification results, allowing researchers and clinicians to detect patterns of misclassification and gain insights into its strengths and weaknesses. Furthermore, the confusion matrix serves as the foundation for calculating other performance metrics including accuracy, sensitivity, specificity, and precision, allowing for a more comprehensive evaluation of the model's performance across classes.

The F1 Score, derived as the harmonic mean of precision and recall, provides a fair assessment of the model's performance, accounting for both false positives and false negatives. The F1 Score generates a single score that indicates the model's overall anticipated accuracy while taking into account the trade-off between precision and recall.

Furthermore, the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is a comprehensive evaluation statistic that assesses the model's discriminatory strength across many classification criteria. The ROC curve provides insights into the model's performance by comparing the True Positive Rate to the False Positive Rate at different threshold settings, whilst the AUC-ROC quantifies the model's overall ability to distinguish between classes.

7.1 Solution Review

According to the results of the evaluation chapter, the ensemble model built for AD classification performs admirably in handling the challenge at hand. The ensemble model has high accuracy, precision, recall, and F1 score, all average 96.67 percent, showing that it can effectively detect cases of Alzheimer's disease while minimising false positives and false negatives. Furthermore, the AUC-ROC score of 0.96 reveals the model's good discriminatory ability to distinguish between positive and negative cases. The extensive evaluation criteria used, such as accuracy, sensitivity, specificity, F1 score, and AUC-ROC, provide a detailed knowledge of the ensemble model's effectiveness in forecasting Alzheimer's disease. These measures emphasise the model's performance across multiple aspects of diagnostic accuracy, providing significant insights for academics and physicians.

Furthermore, the confusion matrix analysis provides precise information about the model's classification results, allowing for the detection of misclassification trends and a more in-depth knowledge of its strengths and limitations. The F1 score, calculated as the harmonic mean of precision and recall, provides a fair assessment of the model's overall performance, taking into account both false positives and false negatives. Furthermore, the AUC-ROC score provides a thorough assessment of the model's discriminatory strength across many categorization criteria.

7.2 Project Review

The project used a comprehensive and systematic process to create an ensemble model for AD classification. We successfully handled many parts of the project, such as data preprocessing, model creation, evaluation, and validation. The model development strategy entailed using advanced deep learning techniques to create and optimise a multimodal ensemble model. Throughout the project, we iteratively updated the model architecture in response to comments from domain experts and evaluation outcomes. This

repeated technique enabled us to properly capture complicated relationship patterns within the data.

The ensemble model's performance and generalizability were assessed using robust validation approaches. Cross-validation techniques were used to assess model performance across many folds of the dataset, ensuring robustness while minimising bias.

Advanced data preprocessing, model creation, validation, and evaluation were all critical talents for our success. In data preprocessing, we developed our skills in managing diverse data sources and identifying relevant information, such as APOE variations, to improve model performance. Our deep learning expertise aided in the construction and optimisation of a multi-modal ensemble model capable of seamlessly integrating multiple data modalities, resulting in comprehensive analysis. Furthermore, our expertise in building rigorous validation techniques allowed us to comprehensively evaluate the model's predictions, assuring their dependability and generalizability across several datasets. These abilities were critical in overcoming obstacles and attaining success on our assignment.

Looking back, if I was to undertake a similar project again, there would be several changes to make. To ensure the durability of the model, improve the model with additional validation techniques like model distillation or adversarial validation. In addition, put additional work into improving the interpretability of the model, maybe by including explainable AI techniques. In conclusion, the project provided useful insights into the application of deep learning in healthcare, emphasising the necessity of interdisciplinary collaboration, continuous learning, and iterative refinement for addressing difficult healthcare concerns.

7.3 Conclusion

The project yielded primary and secondary results in three major areas: background, problem definition, and solution strategy. To begin, the complexities of AD pathology, as well as the involvement of genetic variables such as APOE variations, in diagnosis and prognosis, have been highlighted. Furthermore, the significance of combining several data modalities, such as neuroimaging, CSF biomarkers, and genetic data, to acquire a holistic knowledge of the disease and its course has been emphasised.

Proceeding on to the problem description, many barriers to integrating diverse data sources have been highlighted. Difficulties with preprocessing and feature extraction,

particularly for genetic data such as APOE variations, have been identified. Furthermore, the complexity of model creation, which necessitates capturing subtle disease patterns and linkages within the data to create reliable predictive models for Alzheimer's disease, has been acknowledged.

In terms of solution strategy, it has been determined that specialised preprocessing pipelines and sophisticated deep learning algorithms, when combined with a multi-modal ensemble model framework, provide a viable solution for effectively integrating varied data modalities. Furthermore, the significance of rigorous validation processes, particularly cross-validation, in ensuring the reliability and generalizability of prediction models in clinical practice has been emphasised.

7.4 Future Work

When prioritising project enhancements, interpretability and explainability emerge as the most important criteria. Designing the AI multi-model to produce interpretable outcomes is critical for instilling confidence in its predictions. By shining light on the underlying traits and patterns that influence its decisions, the system can give physicians with useful insights into its decision-making process. Comprehensive visualisations and feature importance rankings will not only improve understanding, but will also help with clinical decision-making, increasing the model's utility in real-world contexts.

Following closely following is the model's ongoing improvement. Refinement and optimisation of model architecture, hyperparameters, and training processes are critical for improving predictive performance and generalizability. Prioritising continuous model improvement guarantees that the system can adapt to changing data patterns while remaining relevant and effective over time, maximising its influence in clinical practice.

While improved data integration is critical for increasing the model's predictive skills, it is prioritised slightly lower in this context. Continued research into ways for seamlessly integrating other data sources, such as demographic information and longitudinal data, is critical, but it can be addressed once the fundamental model architecture and training procedures have been optimised.

Bibliography

[1] Y. Gupta, R. K. Lama, and G.-R. e. a. Kwon, "Prediction and classification of alzheimer's disease based on combined features from apolipoprotein-e genotype, cerebrospinal fluid, mr, and fdg-pet imaging biomarkers," *Frontiers in Computational Neuroscience*, vol. 13, 2019. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fncom.2019.00072

[2] T. M. Ghazal, S. Abbas, S. Munir, M. Khan, M. Ahmad, G. F. Issa, S. B. Zahra, M. A. Khan, and M. K. Hasan, "Alzheimer disease detection empowered with transfer learning." *Computers, Materials & Continua*, vol. 70, no. 3, 2022.

[3] R. Khan, S. Akbar, A. Mehmood, F. Shahid, K. Munir, N. Ilyas, M. Asif, and Z. Zheng, "A transfer learning approach for multiclass classification of alzheimer's disease using mri images," *Frontiers in Neuroscience*, vol. 16, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2022.1050777

[4] A. Society, "The progression, signs and stages of dementia," vol. 1, no. 1, p. 1, February 2021. [Online]. Available: https://www.alzheimers.org.uk/about-dementia/symptoms-and-diagnosis/how-dementia-progresses/progression-stages-dementia

[5] ——, "Alzheimer's disease," vol. 1, no. 1, p. 1, April 2023. [Online]. Available: https://www.alzheimers.org.uk/about-dementia/types-dementia/alzheimers-disease

[6] A. Association, "Alzheimer's disease clinical trials fact sheet," vol. 4, no. 1, pp. 1–4. [Online]. Available: https://www.alz.org/images/northcentraltexas/clinical_trials.pdf

[7] A. Society, "The 10 benefits of early diagnosis," no. 1, p. 1, 2023. [Online]. Available: https://alzheimer.ca/en/about-dementia/do-i-have-dementia/how-get-tested-dementia-tips-individuals-families-friends/10

[8] J. Wu, X.-Y. Chen, H. Zhang, L.-D. Xiong, H. Lei, and S.-H. Deng, "Hyperparameter optimization for machine learning models based on bayesian optimizationb," *Journal of Electronic Science and Technology*, vol. 17, no. 1, pp.

26–40, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1674862X19300047

[9] S. Sivapatham, C. Baskar, S. Anbalagan, S. Gupta, and R. Charan, "Early diagnosis of alzheimer disease using ga optimized transfer learning techniques," in *2022 8th International Conference on Signal Processing and Communication (ICSC)*, 2022, pp. 353–357.

[10] J. Zhang, X. He, Y. Liu, Q. Cai, H. Chen, and L. Qing, "Multi-modal cross-attention network for alzheimer's disease diagnosis with multi-modality data," *Computers in Biology and Medicine*, vol. 162, p. 107050, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482523005152

[11] M. Golovanevsky, C. Eickhoff, and R. Singh, "Multimodal attention-based deep learning for Alzheimer's disease diagnosis," *Journal of the American Medical Informatics Association*, vol. 29, no. 12, pp. 2014–2022, 09 2022. [Online]. Available: https://doi.org/10.1093/jamia/ocac168

[12] S. Roy, R. Mehera, R. K. Pal, and S. K. Bandyopadhyay, "Hyperparameter optimization for deep neuralnetwork models: A comprehensive study onmethods and techniques," 2023.

[13] M. Golovanevsky, C. Eickhoff, and R. Singh, "Multimodal attention-based deep learning for alzheimer's disease diagnosis," *Journal of the American Medical Informatics Association*, vol. 29, no. 12, pp. 2014–2022, 2022.

[14] F. Zhang, M. Petersen, L. Johnson, J. Hall, and S. E. O'Bryant, "Accelerating hyperparameter tuning in machine learning for alzheimer's disease with high performance computing," *Frontiers in Artificial Intelligence*, vol. 4, 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frai.2021.798962

[15] F. Zhang, M. Petersen, L. Johnson, J. Hall, and S. E. O'bryant, "Hyperparameter tuning with high performance computing machine learning for imbalanced alzheimer's disease data," *Applied Sciences*, vol. 12, no. 13, p. 6670, 2022.

[16] S. Venkatasubramanian, J. N. Dwivedi, S. Raja, N. Rajeswari, J. Logeshwaran, A. Praveen Kumar *et al.*, "Prediction of alzheimer's disease using dho-based pretrained cnn model," *Mathematical Problems in Engineering*, vol. 2023, 2023.

[17] A. Khan and S. Zubair, "An improved multi-modal based machine learning approach for the prognosis of alzheimer's disease," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, Part A, pp. 2688–2706, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1319157820303396

[18] R. P. Neelakandan, R. Kandasamy, B. Subbiyan, and M. A. Bennet, "Early detection of alzheimerś disease: An extensive review of advancements in machine learning mechanisms using an ensemble and deep learning technique," *Engineering Proceedings*, vol. 59, no. 1, 2023. [Online]. Available: https://www.mdpi.com/2673-4591/59/1/10

[19] M. Sethi, S. Ahuja, S. Rani, P. Bawa, and A. Zaguia, "Classification of alzheimer's disease using gaussian-based bayesian parameter optimization for deep convolutional lstm network," *Computational and Mathematical Methods in Medicine*, vol. 2021, pp. 1–16, 2021.

[20] A. Thomas, "Comparative study of cnn models for detection of alzheimer's disease."

[21] A. Varghese, B. George, V. Sherimon, A. Shuaily, and H. Salim, "Enhancing trust in alzheimer's disease classification using explainable artificial intelligence: Incorporating local post hoc explanations for a glass-box model." *Bahrain Medical Bulletin*, vol. 45, no. 2, 2023.

[22] J. Zhu, Y. Tan, R. Lin, J. Miao, X. Fan, Y. Zhu, P. Liang, J. Gong, and H. He, "Efficient self-attention mechanism and structural distilling model for alzheimer's disease diagnosis," *Computers in Biology and Medicine*, vol. 147, p. 105737, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482522005108

[23] H. Kang and D.-Y. Kang, "Alzheimer's disease prediction using attention mechanism with dual-phase 18f-florbetaben images," *Nuclear Medicine and Molecular Imaging*, vol. 57, no. 2, pp. 61–72, 2023.

[24] S.-H. Wang, Q. Zhou, M. Yang, and Y.-D. Zhang, "Advian: Alzheimer's disease vgg-inspired attention network based on convolutional block attention module and multiple way data augmentation," *Frontiers in Aging Neuroscience*, vol. 13, 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnagi.2021.687456

[25] X. Sun, W. Guo, and J. Shen, "Toward attention-based learning to predict the risk of brain degeneration with multimodal medical data," *Frontiers in Neuroscience*, vol. 16, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2022.1043626

[26] X. Wu, S. Gao, J. Sun, Y. Zhang, and S. Wang, "Classification of alzheimer's disease based on weakly supervised learning and attention mechanism," *Brain Sciences*, vol. 12, no. 12, 2022. [Online]. Available: https://www.mdpi.com/2076-3425/12/12/1601

[27] F. Liu, H. Wang, Y. Chen, Y. Quan, and L. Tao, "Convolutional neural network based on feature enhancement and attention mechanism for alzheimer's disease prediction using mri images," in *Thirteenth International Conference on Graphics and Image Processing (ICGIP 2021)*, vol. 12083. SPIE, 2022, pp. 281–295.

[28] R. Kadri, B. Bouaziz, M. Tmar, and F. Gargouri, "Depthwise separable convolution resnet with attention mechanism for alzheimer's detection*," in *2022 International Conference on Technology Innovations for Healthcare (ICTIH)*, 2022, pp. 47–52.

[29] B. A. Athena George, L. S. Neetha George, and S. Ramachandran, "An efficient 3d cnn framework with attention mechanisms for alzheimer's disease classification," *Computer Systems Science and Engineering*, vol. 47, no. 2, pp. 2097–2118, 2023. [Online]. Available: http://www.techscience.com/csse/v47n2/53639

[30] D. Jin, J. Xu, K. Zhao, F. Hu, Z. Yang, B. Liu, T. Jiang, and Y. Liu, "Attention-based 3d convolutional network for alzheimer's disease diagnosis and biomarkers exploration," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 1047–1051.

[31] Y. Chen, "An alzheimer's disease identification and classification model based on the convolutional neural network with attention mechanisms." *Traitement du Signal*, vol. 38, no. 5, 2021.

[32] M. Symms, H. R. Jäger, K. Schmierer, and T. A. Yousry, "A review of structural magnetic resonance neuroimaging," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 75, no. 9, pp. 1235–1244, 2004. [Online]. Available: https://jnnp.bmj.com/content/75/9/1235

[33] E. S. Finn, R. A. Poldrack, and J. M. Shine, "Functional neuroimaging as a catalyst for integrated neuroscience," *Nature*, vol. 623, no. 7986, pp. 263–273, 2023.

[34] S. Liu, W. Cai, S. Liu, F. Zhang, M. Fulham, D. Feng, S. Pujol, and R. Kikinis, "Multimodal neuroimaging computing: the workflows, methods, and platforms," *Brain informatics*, vol. 2, no. 3, pp. 181–195, 2015.

[35] Y. Asim, B. Raza, A. K. Malik, S. Rathore, L. Hussain, and M. A. Iftikhar, "A multi-modal, multi-atlas-based approach for alzheimer detection via machine learning," *International Journal of Imaging Systems and Technology*, vol. 28, no. 2, pp. 113–123, 2018. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/ima.22263

[36] M. Amini, M. M. Pedram, A. Moradi, M. Jamshidi, and M. Ouchani, "Single and combined neuroimaging techniques for alzheimer's disease detection," *Computational Intelligence and Neuroscience*, vol. 2021, 2021.

[37] M. Odusami, R. Maskeliūnas, R. Damaševičius, and S. Misra, "Machine learning with multimodal neuroimaging data to classify stages of alzheimer's disease: a systematic review and meta-analysis," *Cognitive Neurodynamics*, pp. 1–20, 2023.

[38] W. Huang, "Multimodal contrastive learning and tabular attention for automated alzheimer's disease prediction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 2473–2482.

[39] L. Liu, S. Zhao, H. Chen, and A. Wang, "A new machine learning method for identifying alzheimer's disease," *Simulation Modelling Practice and Theory*, vol. 99, p. 102023, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1569190X19301546

[40] G. Giuletti, M. Dayan, L. Serra, E. Tuzzi, B. Spano, M. Cercignani, C. Caltagirone, and M. Bozzali, "Alzheimer's disease prediction based on machine learning methods applied to multimodal mr features," in *21st Annual Meeting and Exhibition of the International Society for Magnetic Resonance in Medicine (ISMRM 2013)*, 2013.

[41] H. Cai, X. Huang, Z. Liu, W. Liao, H. Dai, Z. Wu, D. Zhu, H. Ren, Q. Li, T. Liu, and X. Li, "Multimodal approaches for alzheimer's detection using patients' speech and transcript," in *Brain Informatics: 16th International Conference, BI 2023, Hoboken, NJ, USA, August 1–3, 2023, Proceedings*. Berlin, Heidelberg: Springer-Verlag, 2023, p. 395–406. [Online]. Available: https://doi.org/10.1007/978-3-031-43075-6_34

[42] E. Westman, J.-S. Muehlboeck, and A. Simmons, "Combining mri and csf measures for classification of alzheimer's disease and prediction of mild cognitive impairment conversion," *NeuroImage*, vol. 62, no. 1, pp. 229–238, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811912004521

[43] S. Vos, I. van Rossum, L. Burns, D. Knol, P. Scheltens, H. Soininen, L.-O. Wahlund, H. Hampel, M. Tsolaki, L. Minthon, R. Handels, G. L'Italien, W. van der Flier, P. Aalten, C. Teunissen, F. Barkhof, K. Blennow, R. Wolz, D. Rueckert, F. Verhey, and P. J. Visser, "Test sequence of csf and mri biomarkers for prediction of ad in subjects with mci," *Neurobiology of Aging*, vol. 33, no. 10, pp. 2272–2281, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0197458011005495

[44] R. Haulcy and J. Glass, "Classifying alzheimer's disease using audio and text-based representations of speech," *Frontiers in Psychology*, vol. 11, 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.624137

[45] P. N. Young, M. Estarellas, E. Coomans, M. Srikrishna, H. Beaumont, A. Maass, A. V. Venkataraman, R. Lissaman, D. Jiménez, M. J. Betts *et al.*, "Imaging biomarkers in neurodegeneration: current and future practices," *Alzheimer's research & therapy*, vol. 12, no. 1, pp. 1–17, 2020.

[46] G. Lee, K. Nho, B. Kang, K.-A. Sohn, and D. Kim, "Predicting alzheimer's disease progression using multi-modal deep learning approach," *Scientific reports*, vol. 9, no. 1, p. 1952, 2019.

[47] S. Tabarestani, M. Aghili, M. Eslami, M. Cabrerizo, A. Barreto, N. Rishe, R. E. Curiel, D. Loewenstein, R. Duara, and M. Adjouadi, "A distributed multitask multimodal approach for the prediction of alzheimer's disease in a longitudinal study," *NeuroImage*, vol. 206, p. 116317, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811919309085

[48] S. H. Hojjati, A. Babajani-Feremi, and t. A. D. N. I. , "Prediction and modeling of neuropsychological scores in alzheimer's disease using multimodal neuroimaging data and artificial neural networks," *Frontiers in Computational Neuroscience*, vol. 15, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fncom.2021.769982

[49] M. Ten Kate, A. Redolfi, E. Peira, I. Bos, S. J. Vos, R. Vandenberghe, S. Gabel, J. Schaeverbeke, P. Scheltens, O. Blin *et al.*, "Mri predictors of amyloid pathology: results from the emif-ad multimodal biomarker discovery study," *Alzheimer's research & therapy*, vol. 10, no. 1, pp. 1–12, 2018.

[50] M. F. Romano, A. Balachandra, X. Zhou, M. Jadick, S. Qiu, D. Nijhawan, S. P. Chin, R. Au, and V. B. Kolachalama, "Comparative analysis of cerebrospinal fluid markers and multimodal imaging in predicting alzheimer's disease progression," *Alzheimer's & Dementia*, vol. 17, no. S4, p. e054457, 2021. [Online]. Available: https://alz-journals.onlinelibrary.wiley.com/doi/abs/10.1002/alz.054457

[51] U. Khatri and G.-R. Kwon, "An efficient combination among smri, csf, cognitive score, and apoe ε4 biomarkers for classification of ad and mci using extreme learning machine," *Computational Intelligence and Neuroscience*, vol. 2020, 2020.

[52] H. Jang, J. Park, S. Woo, S. Kim, H. J. Kim, D. L. Na, S. N. Lockhart, Y. Kim, K. W. Kim, S. H. Cho, S. J. Kim, J.-K. Seong, and S. W. Seo, "Prediction of fast decline in amyloid positive mild cognitive impairment patients using multimodal biomarkers," *NeuroImage: Clinical*, vol. 24, p. 101941, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2213158219302918

[53] A. Thushara, C. UshaDevi Amma, A. John, and R. Saju, "Multimodal mri based classification and prediction of alzheimer's disease using random forest ensemble," in *2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA)*, 2020, pp. 249–256.

[54] R. Kumari, A. Nigam, and S. Pushkar, "An efficient combination of quadruple biomarkers in binary classification using ensemble machine learning technique for early onset of alzheimer disease," *Neural Computing and Applications*, vol. 34, no. 14, pp. 11 865–11 884, 2022.

[55] J. D. V. Oriol, E. E. Vallejo, K. Estrada, J. G. T. Peña, A. D. N. Initiative *et al.*, "Benchmarking machine learning models for late-onset alzheimer's disease prediction from genomic data," *BMC bioinformatics*, vol. 20, 2019.

[56] A. Ezzati, A. R. Zammit, D. J. Harvey, C. Habeck, C. B. Hall, R. B. Lipton, A. D. N. Initiative *et al.*, "Optimizing machine learning methods to improve predictive models of alzheimer's disease," *Journal of Alzheimer's Disease*, vol. 71, no. 3, pp. 1027–1036, 2019.

[57] S. Gaubert, M. Houot, F. Raimondo, M. Ansart, M.-C. Corsi, L. Naccache, J. D. Sitt, M.-O. Habert, B. Dubois, F. De Vico Fallani, S. Durrleman, and S. Epelbaum, "A machine learning approach to screen for preclinical alzheimer's disease," *Neurobiology of Aging*, vol. 105, pp. 205–216, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0197458021001470

[58] X. Feng, F. A. Provenzano, S. A. Small, and A. D. N. Initiative, "A deep learning mri approach outperforms other biomarkers of prodromal alzheimer's disease," *Alzheimer's Research & Therapy*, vol. 14, no. 1, p. 45, 2022.

[59] A. Loddo, S. Buttau, and C. Di Ruberto, "Deep learning based pipelines for alzheimer's disease diagnosis: A comparative study and a novel deep-ensemble method," *Computers in Biology and Medicine*, vol. 141, p. 105032, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S001048252100826X

[60] A. Alorf and M. U. G. Khan, "Multi-label classification of alzheimer's disease stages from resting-state fmri-based correlation connectivity data and deep learning," *Computers in Biology and Medicine*, vol. 151, p. 106240, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482522009489

[61] R. Sampath, M. Baskar, and V. B. Devi, "Hippocampus region's volume-based alzheimer's stages detection using a deep learning model," in *2022 1st International Conference on Computational Science and Technology (ICCST)*, 2022, pp. 74–79.

[62] Lee, Garam, Nho, and K. et al, "Predicting alzheimer's disease progression using multi-modal deep learning approach," *Scientific Reports*, vol. 9, no. 1, p. 1952, Feb 2019. [Online]. Available: https://doi.org/10.1038/s41598-018-37769-z

[63] H. Choi and K. H. Jin, "Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging," *Behavioural Brain Research*, vol. 344, pp. 103–109, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0166432818301013

[64] S. Maddury and K. Desai, "Deepad: A deep learning application for predicting amyloid standardized uptake value ratio through pet for alzheimer's prognosis," *Frontiers in Artificial Intelligence*, vol. 6, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frai.2023.1091506

[65] M. Yildirim and A. Çinar, "Classification of alzheimer's disease mri images with cnn based hybrid method," *Ingénierie des systèmes d information*, vol. 25, pp. 413–418, 09 2020.

[66] G. Battineni, N. Chintalapudi, F. Amenta, and E. Traini, "A comprehensive machine-learning model applied to magnetic resonance imaging (mri) to predict alzheimer's disease (ad) in older subjects," *Journal of Clinical Medicine*, vol. 9, no. 7, 2020. [Online]. Available: https://www.mdpi.com/2077-0383/9/7/2146

[67] Tushar, R. K. Patel, E. Aggarwal, K. Solanki, O. Dahiya, and S. A. Yadav, "A logistic regression and decision tree based hybrid approach to predict alzheimer's disease," in *2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)*, 2023, pp. 722–726.

[68] M. S. Kumar, H. Azath, A. K. Velmurugan, K. Padmanaban, and M. Subbiah, "Prediction of Alzheimer's disease using hybrid machine learning technique," *AIP Conference Proceedings*, vol. 2523, no. 1, p. 020091, 01 2023. [Online]. Available: https://doi.org/10.1063/5.0110283

[69] A. Khan and S. Zubair, "Development of a three tiered cognitive hybrid machine learning algorithm for effective diagnosis of alzheimer's disease," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 10, Part A, pp. 8000–8018, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1319157822002439

[70] G. S. Babu, S. Rao, and R. R. Rao, "Alzheimer's disease prediction by hybrid cnn and svm classifier with metaheuristic approach," *International Journal of Computing and Digital Systems*, vol. 12, no. 1, pp. 1119–1128, 2022.

[71] M. Shcherbakov and C. Sai, "A hybrid deep learning framework for intelligent predictive maintenance of cyber-physical systems," *ACM Transactions on Cyber-Physical Systems*, 05 2022.

[72] T. Abuhmed, S. El-Sappagh, and J. M. Alonso, "Robust hybrid deep learning models for alzheimer's progression detection," *Knowledge-Based Systems*, vol. 213, p. 106688, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705120308170

[73] H. Saleh, N. ElRashidy, M. Abd Elaziz, A. O. Aseeri, and S. El-Sappagh, "Genetic algorithms based optimized hybrid deep learning model for explainable alzheimer's prediction based on temporal multimodal cognitive data," 2023.

[74] N. Mahendran, P. M. D. R. Vincent, K. Srinivasan, and C.-Y. Chang, "Improving the classification of alzheimer's disease using hybrid gene selection pipeline and deep learning," *Frontiers in Genetics*, vol. 12, 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fgene.2021.784814

[75] S. Liu, Y. Zheng, H. Li, M. Pan, Z. Fang, M. Liu, Y. Qiao, N. Pan, W. Jia, and X. Ge, "Improving alzheimer diagnoses with an interpretable deep learning framework: Including neuropsychiatric symptoms," *Neuroscience*, vol. 531, pp. 86–98, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306452223004116

[76] A. Sarath Vignesh, H. Denicke Solomon, P. Dheepan, and G. Kavitha, "Segmentation and severity classification of dementia in magnetic resonance imaging using deep learning networks," in *2023 International Conference on Bio Signals, Images, and Instrumentation (ICBSII)*, 2023, pp. 1–6.

[77] P. Moore, T. Lyons, J. Gallacher, and A. D. N. Initiative, "Random forest prediction of alzheimer's disease using pairwise selection from time series data," *PloS one*, vol. 14, no. 2, p. e0211558, 2019.

[78] J. Kim, S. Kim, D. Kang, S. Kim, R. Kwon, D. Yon, and J. Kim, "Feature extraction of time series data on functional near-infrared spectroscopy and comparison of deep learning performance for classifying patients with alzheimer's-related mild cognitive impairment: a post-hoc analysis of a diagnostic interventional trial," *Eur Rev Med Pharmacol Sci*, vol. 27, no. 14, pp. 6824–6830, 2023.

[79] N. Rahim, S. El-Sappagh, S. Ali, K. Muhammad, J. Del Ser, and T. Abuhmed, "Prediction of alzheimer's progression based on multimodal deep-learning-based fusion and visual explainability of time-series data,"

Information Fusion, vol. 92, pp. 363–388, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253522002378

[80] S. El-Sappagh, T. Abuhmed, S. Riazul Islam, and K. S. Kwak, "Multimodal multitask deep learning model for alzheimer's disease progression detection based on time series data," *Neurocomputing*, vol. 412, pp. 197–215, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231220309383

[81] N. Rahim, T. Abuhmed, S. Mirjalili, S. El-Sappagh, and K. Muhammad, "Time-series visual explainability for alzheimer's disease progression detection for smart healthcare," *Alexandria Engineering Journal*, vol. 82, pp. 484–502, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1110016823008384

[82] S. El-Sappagh, F. Ali, T. Abuhmed, J. Singh, and J. M. Alonso, "Automatic detection of alzheimer's disease progression: An efficient information fusion approach with heterogeneous ensemble classifiers," *Neurocomputing*, vol. 512, pp. 203–224, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231222010955

[83] 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. 72 726–72 735, 2019.

[84] M. Grassi, D. A. Loewenstein, D. Caldirola, K. Schruers, R. Duara, and G. Perna, "A clinically-translatable machine learning algorithm for the prediction of alzheimer's disease conversion: further evidence of its accuracy via a transfer learning approach," *International Psychogeriatrics*, vol. 31, no. 7, p. 937–945, 2019.

[85] A. Danker and J. Wirgård Wiklund, "Using transfer learning to classify different stages of alzheimer's disease," 2021.

[86] "Icbm template." [Online]. Available: http://www.bmap.ucla.edu/portfolio/atlases/ICBM_Template/

[87] S.-Q. Wang, Z. Zhang, F. He, and Y. Hu, "Editorial: Generative ai for brain imaging and brain network construction," *Frontiers in Neuroscience*, vol. 17, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2023.1279470

[88] Y. Kimura, A. Watanabe, T. Yamada, S. Watanabe, T. Nagaoka, M. Nemoto, K. Miyazaki, K. Hanaoka, H. Kaida, and K. Ishii, "Ai approach of cycle-consistent generative adversarial networks to synthesize pet images to train computer-aided diagnosis algorithm for dementia," *Annals of Nuclear Medicine*, vol. 34, pp. 512–515, 2020.

[89] R. Wang, V. Bashyam, Z. Yang, F. Yu, V. Tassopoulou, S. S. Chintapalli, I. Skampardoni, L. P. Sreepada, D. Sahoo, K. Nikita, A. Abdulkadir, J. Wen, and C. Davatzikos, "Applications of generative adversarial networks in neuroimaging and clinical neuroscience," *NeuroImage*, vol. 269, p. 119898, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811923000472

[90] C. Han, L. Rundo, K. Murao, Z. Á. Milacski, K. Umemoto, E. Sala, H. Nakayama, and S. Satoh, "Gan-based multiple adjacent brain mri slice reconstruction for unsupervised alzheimer's disease diagnosis," in *Computational Intelligence Methods for Bioinformatics and Biostatistics*, P. Cazzaniga, D. Besozzi, I. Merelli, and L. Manzoni, Eds. Cham: Springer International Publishing, 2020, pp. 44–54.

[91] D. F. Santos, "Brain tumor detection using the vgg-16 model: A deep learning aproach," *Research Gate*, p. 1, 2023.

[92] scikit learn, "sklearn.preprocessing.standardscaler," p. 1.

[93] I. Dabbura, "K-means clustering: Algorithm, applications, evaluation methods, and drawbacks," *Medium*, p. 1, 2018.

[94] M. Kalirane, "Ensemble learning in machine learing: Bagging, boosting and stacking," *Analytics Vidhya*, p. 1, 2024.

[95] J. Brownlee, "A gentle introduction to mini-batch gradient descent and how to configure batch size," *Machine Learning Mastery*, p. 1, 2019.