Towards Accessible Healthcare: Machine Learning-Enabled Diagnosis of Alzheimer's Disease

by

Asif Rasheed

A THESIS
SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE
AND THE FACULTY OF GRADUATE STUDIES
OF LAKEHEAD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE (SPECIALIZATION IN ARTIFICIAL INTELLIGENCE)

Supervisory Committee

Dr. Zubair Fadlullah
Supervisor
(*External Adjunct Professor, Department of Computer Science, Lakehead University, Thunder Bay, Ontario, Canada.*
Associate Professor, Department of Computer Science, University of Western Ontario, London, Ontario, Canada.)

Dr. Mostafa Fouda
Co-Supervisor
(*Associate Professor, Department of Electrical and Computer Engineering, Idaho State University, Pocatello, Idaho, USA.*)

Dr. Garima Bajwa
Internal Examiner
(*Assistant Professor, Department of Computer Science, Lakehead University, Thunder Bay, Ontario, Canada.*)

Dr. Mohamed Ibrahem
External Examiner
(*Assistant Professor, School of Computer and Cyber Sciences, Augusta University, Augusta, Georgia, USA.*)

ABSTRACT

Alzheimer's disease poses a critical challenge to public health with an increasing prevalence among the aging population worldwide. The research question is whether machine learning-based solutions could be a reliable, cost-effective, and non-invasive alternative to existing biomarker tests. This thesis presents two machine learning-based approaches to diagnosing Alzheimer's disease using Magnetic Resonance Imaging (MRI) and blood-based biomarkers. The first approach aims to train machine learning models on volumes of brain regions from MRI to classify patients into three classes: Alzheimer's Dementia (AD), Mild Cognitive impairment (MCI) and Normal Control (NL). Pretrained weights of a well-known CNN-based brain segmentation model were used in segmenting the hippocampal, parahippocampal, ventricles, entorhinal and cerebral white matter from MRI of patients, and their volumes were estimated. The volumes and demographic data of the patients were subsequently trained on SVM and KNN models, and their performance was recorded.

The second approach aims to design efficient feature selection methods to identify relevant feature panels to identify individuals in the early stages of Alzheimer's accurately. Two feature selection methods were introduced. The first method ranks features according to their dependence on the diagnosis, determined using metrics such as Mutual Information, Symmetric Uncertainty and Cramer's V. Panels are formed in this method by iteratively selecting the top features and increasing the panel size. The second method filters out irrelevant features using the Euclidean distance between the class means of each feature and applying a threshold. Candidate panels identified using the two methods are extensively tested on two datasets, and their performances are reported. The blood-based approach is further expanded to allow multiclass classification of patients into three classes: AD, MCI and NL. While the first feature selection method remains largely unchanged, three new scoring techniques are used in the Euclidean distance-based approach to allow this. These approaches include scoring features using the minimum, maximum and average Euclidean distances between the class means for each feature. All possible combinations of relevant features identified using the two methods are extensively tested using two datasets, and their performances are reported.

Finally, we surveyed previous works demonstrating the significance of near-infrared as a non-invasive and accessible tool for diagnosing Alzheimer's disease. In conclusion, the two approaches introduced in this thesis offer efficient and accurate diagnosis of Alzheimer's disease. The first approach achieved an accuracy of 99.74%, and the second achieved an accuracy of 99.06% in binary classification and 91.96% in multiclass classification. The performances prove that the approaches are viable as reliable, cost-effective and relatively non-invasive alternatives to traditional biomarker tests, facilitating accessible diagnosis of Alzheimer's disease.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Dr Zubair Fadlullah, and my co-supervisor, Dr Mostafa Fouda, for all their support and guidance during this research. Their guidance allowed me to concentrate on the research even in spite of personal challenges, and their insights were extremely helpful in determining the direction of the research. Furthermore, their expert advice, patience, and constructive feedback were extremely valuable in allowing me to successfully complete this research on time.

I would also like to acknowledge and express my gratitude for the support and encouragement of my family and friends. My family's unwavering belief in me and emotional support were invaluable throughout this journey. The friendship and companionship of my friends made me feel home away from home. Their unwavering support and camaraderie provided comfort during challenging times. This thesis not only represents an academic achievement but also serves as a product of the support from all those who stood by me along the way.

Finally, I would like to express my gratitude to Lakehead University and the Faculty of Graduate Studies for giving me the opportunity and academic environment to conduct this research.

PUBLICATIONS

Parts of this thesis have been submitted for peer-review, published or accepted for publication:

- **Optimizing Feature Panels for Effective Model Training Toward Early Diagnosis of Alzheimer's Disease** has been accepted in the 2024 International Conference on Smart Applications, Communications and Networking (SmartNets). (part of Chapter 4)

Contents

Supervisory Committee ii

Abstract iii

Acknowledgements iv

Publications v

Table of Contents vi

List of Tables ix

List of Figures x

List of Algorithms xi

1 Introduction 1
- 1.1 Contributions . . . 3
- 1.2 Thesis Structure . . . 3

2 Background 4
- 2.1 Alzheimer's Disease . . . 5
  - 2.1.1 Alzheimer's Diagnosis . . . 5
- 2.2 Machine Learning . . . 6
  - 2.2.1 Supervised Learning . . . 6
    - 2.2.1.1 K-Nearest Neighbor . . . 7
    - 2.2.1.2 Support Vector Machine . . . 7
    - 2.2.1.3 Neural Networks . . . 8
    - 2.2.1.4 Hyperparameter Optimization . . . 8
    - 2.2.1.5 Class-Imbalance Problem . . . 9
  - 2.2.2 Performance Evaluation of ML Models . . . 9
    - 2.2.2.1 Cross-Validation . . . 9
    - 2.2.2.2 Accuracy . . . 9

2.2.2.3 Recall . . . 10
2.2.2.4 F1-Score . . . 10
2.2.2.5 Statistical Significance . . . 10
2.3 Information Measures . . . 11
2.3.0.1 Correlation . . . 11
2.3.0.2 Mutual Information . . . 11
2.3.0.3 Cramer's V . . . 12
2.4 Tools and Approach . . . 12

3 Alzheimer's Disease Stage Classification using MRI 13
3.1 Introduction . . . 14
3.2 Related Works . . . 15
3.3 Problem Description . . . 16
3.4 Dataset Preparation . . . 16
3.5 Proposed Methodology . . . 17
3.6 Performance Evaluation . . . 19

4 Identifying Biomarker Panels for Early Detection of Alzheimer's Disease using Blood 20
4.1 Introduction . . . 21
4.2 Related Work . . . 23
4.3 Problem Description . . . 24
4.4 Dataset Preparation . . . 25
4.5 Proposed Feature Selection Methods . . . 25
4.5.1 Dependency-Dependent Feature Selection Method . . . 25
4.5.2 Euclidean Distance-based Feature Selection Method . . . 27
4.6 Performance Evaluation . . . 30

5 Identifying Biomarker Panels for Alzheimer's Disease Stage Classification using Blood 37
5.1 Introduction . . . 38
5.2 Related Work . . . 39
5.3 Problem Statement . . . 40
5.4 Methods . . . 40
5.4.1 Dataset Preparation . . . 40
5.4.2 Feature Selection . . . 41
5.4.2.1 Minimum Distance Approach . . . 41
5.4.2.2 Maximum Distance Approach . . . 41
5.4.2.3 Average Distance Approach . . . 42

5.4.3 Training and Validation 42
5.5 Performance Evaluation 42

6 Future Scope: Near Infrared-Based Disease Detection as a Non-Invasive Approach 50

7 Conclusions 52
7.1 Limitations 53

Bibliography 54

List of Tables

Table 3.1 Brainchop Model Architecture 17
Table 3.2 Performance Comparison for MRI-based methods 19

Table 4.1 Optimal Panels Identified in Previous Work 22
Table 4.2 Performance Comparsion for Blood-based Methods 32

Table 5.1 Performance Comparsion for Blood-based Multiclass Methods 45

List of Figures

Figure 3.1 MRI before and after segmentation 16
Figure 3.2 Overall Framework for Disease Stage Classification using MRI. . . . 18

Figure 4.1 Overall Framework for Dependency-Dependent Feature Selection. . . 27
Figure 4.2 Feature space, f_3f_2-plane, f_1f_2-plane and f_1f_3-plane (clockwise). . . 28
Figure 4.3 Class means for each feature. 29
Figure 4.4 Overall Framework for Euclidean Distance-based Feature Selection. . 29
Figure 4.5 SN (sensitivity), SP (specificity), and accuracy (ACC) for Set 1 during testing (dependency-dependent method). 33
Figure 4.6 SN, SP, and accuracy for Set 2 during testing (dependency-dependent method). 34
Figure 4.7 SN, SP, and accuracy for Set 1 during testing (Euclidean-Distance method). 35
Figure 4.8 SN, SP, and accuracy for set 2 during testing (Euclidean-Distance method). 36
Figure 4.9 Testing Accuracy for different threshold values. 36

Figure 5.1 Overall Framework for Disease Stage Classification using Blood. . . . 43
Figure 5.2 Recall and Accuracy for Set 1 (Dependency-Dependent method). . . 46
Figure 5.3 Recall and Accuracy for Set 2 (Dependency-Dependent method). . . 47
Figure 5.4 Recall and Accuracy for Set 1 (Euclidean Distance-based Method). . 48
Figure 5.5 Recall and Accuracy for Set 2 (Euclidean Distance-based Method). . 49

List of Algorithms

Algorithm 4.1 Multi-Interval Discretization 26

Chapter 1

Introduction

Dementia is a syndrome arising from nerve cell damage in the brain, causing affected individuals to struggle with memory, language, problem-solving and other cognitive functions that prevent them from carrying out everyday tasks [1]. Alzheimer's is a neurodegenerative disease and the most common cause of dementia [1]. Over 5.7 million individuals across different age groups are affected by Alzheimer's today [1]. The number of affected individuals is expected to rise to 7.1 million by 2025 and up to 13.8 million by 2030 [1]. Many patients who would meet the criteria for an Alzheimer's diagnosis are underdiagnosed by physicians [1]. Furthermore, only a few patients with dementia documented in their medical records in the United States report being informed of their diagnosis [1]. Although medications to prevent or reverse Alzheimer's do not exist today, timely diagnosis is essential to provide necessary support and resources for individuals with dementia [1]. Conversely, an early diagnosis offers patients time for financial and emotional preparedness [1]. Estimates suggest savings up to $7.9 trillion for the US population with accurate early diagnosis of Alzheimer's [1]. Moreover, early diagnosis offers patients the opportunity to participate in experimental treatments [1]. Therefore, there is a need for accurate and accessible diagnostic tools for Alzheimer's [1].

Studies have shown that changes in the brain due to Alzheimer's could manifest as early as 20 years before symptoms are apparent [1]. Individuals' disease stages are classified into Mild Cognitive Impairment (MCI) or dementia due to Alzheimer's (AD), depending on the extent of damage [1]. Individuals with MCI experience symptoms such as behavioural changes or cognitive difficulties that are apparent to them or others they frequently interact with [1]. However, they can still manage their daily activities independently [1]. While the rate at which MCI progresses to AD varies among individuals, studies have shown that 32% of patients with MCI progress to AD in five-year follow-ups [1]. Moreover, individuals with MCI have a higher likelihood of developing dementia compared to healthy individuals [1]. Although some studies have further classified patients into early and late MCI stages [2], for

the purpose of this thesis, individuals will be classified as Normal Control (NL), MCI and AD. Many researchers have succeeded in developing highly accurate Machine Learning (ML) models to classify between NL and AD individuals or NL and MCI individuals; however, classifying between MCI and AD, or among the three classes, remains a challenge [3].

Traditionally, Alzheimer's disease is diagnosed using the means of medical history assessments, cognitive testing using standardized tests and brain imaging using Magnetic Resonance Imaging (MRI) to identify abnormalities in the brain [4]. Today, biomarker tests are considered a critical tool for diagnosing Alzheimer's [5]. Beta-amyloid imaging using Positron Emission Tomography (PET) and Cerebrospinal Fluid (CSF) test are two biomarker tests used today to diagnose Alzheimer's disease [1]. While both CSF tests and PET imaging are highly accurate in identifying beta-amyloid depositions in the brain, there are significant barriers to their accessibility. CSF testing is considered invasive because it requires a lumbar puncture to obtain patient samples, while PET scans are expensive compared to CSF testing [5]. Moreover, the application of CSF testing outside a clinical setting is limited due to the specialized requirements for the test [5]. These factors can limit the widespread use of these tests.

ML models have emerged as powerful tools for diagnosing Alzheimer's disease with high accuracy. Often, these models are trained on MRI and blood data. MRI is favoured for its non-invasive nature. CNN-based architectures such as GoogleNet and ResNet have demonstrated high accuracy in Alzheimer's diagnosis when trained on MRI [6, 7]. Previous works have also used brain atlases for segmenting specific brain regions from MRI to train ML models [8–10]. In contrast, blood tests offer a less expensive and non-invasive alternative to imaging and CSF tests [11]. Previous works using blood have predominantly focused on identifying biomarker panels from blood proteomic data, leading to promising outcomes in the early detection of Alzheimer's disease.

While CNN-based models trained on MRI have achieved high accuracy, their training requires substantial computation resources, often requiring dedicated hardware. Additionally, brain atlases may not be suitable for MRI segmentation in Alzheimer's diagnosis due to the potential abnormal changes in the brain associated with the disease [12]. Moreover, efficient and reliable feature selection methods are needed to identify blood biomarker panels as reproducing performance of previously reported biomarker panels has been unreliable. Furthermore, there is a need to extend classification using the biomarker panels beyond distinguishing Alzheimer's patients from normal individuals to identifying the specific disease stage of patients. The primary objective of this thesis is to address these critical challenges in Alzheimer's diagnosis by developing reliable, cost-effective and non-invasive ML models using MRI and blood data. By addressing the challenges, this thesis aims to contribute to developing accessible diagnostic tools for Alzheimer's disease.

1.1 Contributions

The main contributions of this thesis are:

- A feature extraction approach for MRI data using volumetric analysis of specific brain regions associated with disease progression.
- A feature selection method using dependency between features and disease stages to identify relevant blood protein panels for early detection and disease stage classification of Alzheimer's disease.
- A feature selection method using Euclidean distance between class means identifying relevant blood protein panels for early detection and disease stage classification of Alzheimer's disease.

1.2 Thesis Structure

This thesis is divided into seven chapters. Chapter 1 introduces the topic, the research motivation, and important contributions and outlines the structure of the thesis. Chapter 2 discusses the fundamental concepts associated with the work. In Chapter 3, we explore Alzheimer's disease stage classification using ML models trained on estimated volumes of brain regions derived from MRI scans. In Chapter 4, we introduce two feature selection methods to identify relevant blood protein panels that could help in the early diagnosis of Alzheimer's using ML. Chapter 5 expands the scope of Chapter 4 by exploring the performance of ML models in disease stage classification when trained on protein panels formed using the two feature selection methods. Chapter 6 discusses the current works using Near Infrared (NIR) in diagnosing Alzheimer's, highlighting its significance as a future direction. Finally, Chapter 7 summarizes the contributions of this work and its limitations and concludes the thesis.

Chapter 2

Background

2.1 Alzheimer's Disease . . . 5
2.1.1 Alzheimer's Diagnosis . . . 5
2.2 Machine Learning . . . 6
2.2.1 Supervised Learning . . . 6
2.2.1.1 K-Nearest Neighbor . . . 7
2.2.1.2 Support Vector Machine . . . 7
2.2.1.3 Neural Networks . . . 8
2.2.1.4 Hyperparameter Optimization . . . 8
2.2.1.5 Class-Imbalance Problem . . . 9
2.2.2 Performance Evaluation of ML Models . . . 9
2.2.2.1 Cross-Validation . . . 9
2.2.2.2 Accuracy . . . 9
2.2.2.3 Recall . . . 10
2.2.2.4 F1-Score . . . 10
2.2.2.5 Statistical Significance . . . 10
2.3 Information Measures . . . 11
2.3.0.1 Correlation . . . 11
2.3.0.2 Mutual Information . . . 11
2.3.0.3 Cramer's V . . . 12
2.4 Tools and Approach . . . 12

2.1 Alzheimer's Disease

Alzheimer's is a neurodegenerative disease that gradually progresses over time [1]. The primary change in the brain due to Alzheimer's is the deposition of beta-amyloid proteins around the neurons that interfere with neuron-to-neuron communication [1]. These proteins trigger microglia, the immune cells in the brain, which attempt to remove the beta-amyloid proteins [1]. However, as the severity of the disease progresses, the microglia might not be effective in removing the proteins [1]. With the beta-amyloid deposition worsening, tau proteins spread throught the brain [1]. These proteins interrupt the transport of essential nutrients inside the neurons [1]. The brain shrinks due to the loss of neurons, and the brain functions compromise as the brain's ability to metabolize glucose deteriorates [1]. Besides the abnormal changes in the brain, symptoms of Alzheimer's disease include cognitive difficulties such as memory loss, behavioural changes such as the patient getting aggressive and psychotic manifestations such as hallucinations [13]. While there is no specific rate at which changes in the brain due to Alzheimer's could progress, studies have shown that the changes could begin as early as 20 years before the symptoms manifest in patients [1]. Some methods used in diagnosing Alzheimer's disease are discussed in the following section.

2.1.1 Alzheimer's Diagnosis

Clinical examinations, medical history assessments, cognitive testing and brain imaging are some methods traditionally used in diagnosing Alzheimer's disease [4]. Clinical examinations involve a physician assessing the patient's cognitive skills, such as attention, language, memory and visuospatial skills [4]. Cognitive testing uses standardized examinations such as the Mini-Mental State Examination (MMSE) and the Montreal Cognitive Assessment (MoCA) to determine the amount of cognitive impairment [4]. A medical history assessment helps in ruling out the possibility that the patient's current or past health condition or current medications are causing the symptoms [4]. During this, the history of the patient's closest family members is also assessed [4]. Brain imaging using MRI helps in determining the amount of abnormal changes in the brain due to Alzheimer's disease, such as abnormal rate of atrophy and changes to the brain structure [4].

In 2018, the National Institute on Aging - Alzheimer's Association (NIA-AA) proposed a framework to standardize biomarker testing for Alzheimer's disease [5]. Biomarkers are measurable biological units that help determine the presence, absence or risk of developing a disease [1]. Today, many researches are underway to develop easy and cost-effective diagnoses using biomarkers [1]. Biomarker tests such as beta-amyloid imaging using PET and CSF analysis are used in many regions to aid in Alzheimer's diagnosis [1].

Beta-amyloid imaging using PET is a widely used non-invasive biomarker test used in diagnosing Alzheimer's disease [5]. The test quantifies the amount of beta-amyloid and

tau protein depositions in the brain [5]. The test uses specialized chemical trackers, called radiotracers, injected into the patients [5]. The chemicals bind onto the beta-amyloid and tau proteins in the brain and emit positrons that the PET scanners can detect [5]. While the test could accurately quantify the amount of beta-amyloid deposition, it cannot conclusively diagnose Alzheimer's as high levels of depositions are normal among the elderly population, or it could be due to other neurodegenerative disorders [5].

CSF tests are considered superior compared to beta-amyloid imaging using PET in accurate and timely diagnosis of Alzheimer's disease [14]. CSF samples are obtained from subjects using lumbar puncture by a specialist and analyzed to measure biomarkers [14]. Clinical studies comparing CSF tests with PET imaging showed that in 90% of cases, beta-amyloid presence detected using CSF tests correlated with its presence in PET imaging [14]. However, there were cases where beta-amyloid presence was detected in PET imaging despite the CSF test suggesting its absence [14]. The reliability of CSF tests was significantly improved by using ratios of CSF biomarkers measurements instead of individual biomarkers, suggesting that CSF tests serve as a reliable, cost-effective alternative to beta-amyloid imaging using PET [14]. However, CSF tests are limited to clinical studies due to their specialized requirements [14]. Moreover, the lumbar puncture procedure necessary for these tests is considered invasive [11].

2.2 Machine Learning

Machine learning created a new paradigm in problem-solving using computers by developing algorithms and statistical models that could help in solving problems without requiring explicitly programmed instructions. ML models instead solve problems by learning from datasets. The dependence on data allows them to be used in various applications where appropriate datasets are available. Moreover, the performance of ML models can be continuously improved by using newer data. For example, the performance of a medical diagnosis model trained on data up to 2022 could be improved by learning new data collected between 2022 and 2024.

2.2.1 Supervised Learning

In this thesis, we specifically use supervised learning. Supervised learning is a type of machine learning where the models are taught to map inputs into specific targets. These targets can be continuous values where the models are trained by regression to fit the statistical model of the data. In other cases, the targets can be discrete categories, such as disease stages or binary values, such as positive or negative. Algorithms such as Support Vector Machine (SVM) and K-nearest neighbour (KNN) algorithms deal with such classification problems.

2.2.1.1 K-Nearest Neighbor

K-Nearest Neighbor (KNN) [15] algorithm helps to classify a given input according to the majority in its neighbourbood. For a given unknown point and k, the algorithm determines the distance between the point and all the samples in the training dataset and chooses the k closest samples. The majority label among the neighbours is then identified and assigned to the unknown point. Euclidean distance is commonly used to determine the closest neighbours. However, other distances, such as the Manhattan distance and Minkowski distance, have also been used in this algorithm.

2.2.1.2 Support Vector Machine

Support Vector Machine (SVM) [16] uses the training dataset to fit a hyperplane that could separate the samples according to their classes. There are two main concepts associated with SVMs: margin and support vectors. Margin is the minimum distance between the hyperplane or decision boundary and the samples, and the closest samples are the support vectors. During training, the model tries to fit the best hyperplane that maximizes the margin while accurately separating the samples. Unknown points are classified by finding which side of the hyperplane they belong to.

A simple hyperplane might not be sufficient for complex datasets to separate the samples. In this case, the samples can be transformed into higher-dimensional spaces where a hyperplane could separate them. In practice, we apply kernel functions to treat the samples as if they were in a higher dimension without transforming them by estimating the similarities between each sample using their dot products in a higher dimensional space. Subsequently, The decision boundary becomes non-linear when kernel functions are used. The polynomial and Radial Basis Function (RBF) kernels are used in this thesis. Polynomial kernels use polynomial functions to compute the similarity between the samples, while RBF kernels measure the similarity between samples based on their distance in the feature space.

SVMs separate the samples using a single hyperplane. However, when dealing with multiple classes, one hyperplane might not be sufficient to separate all the classes. One vs All and One vs One approaches are used in such cases. In the One-vs-All approach, n classifiers are trained for n classes. Each classifier determines whether the unknown sample belongs to the class corresponding to that classifier. These classifiers output probability scores and the class associated with the model having the highest probability score is assigned to the unknown sample. Conversely, the One-vs-One approach involves training $\frac{n(n-1)}{2}$ classifiers. Each classifier is dedicated to distinguishing between a unique pair of classes. When classifying an unknown sample, the class is assigned based on the majority output from the classifiers. The one-vs-one approach is used by the Python library used for implementing

the methodologies discussed in this thesis [17].

2.2.1.3 Neural Networks

Neural networks [18] are ML models inspired by the neural networks in the human brain. The fundamental unit of neural network models is the perceptron. Perceptrons aim to identify the optimal hyperplane that separates samples based on their class in a feature space. To achieve this, perceptrons start with a random hyperplane. Then, a random point from the dataset is selected, and its placement relative to the hyperplane is determined. If misplaced, the perceptron estimates the error and updates the hyperplane according to the error and a predefined learning rate. Similar to SVMs, the perceptrons might not be effective for datasets that aren't linearly seperable. Neural networks address the limitation by using networks of layers of perceptrons. Each layer processes the input data through a series of weighted connections and applies an activation function to produce an output that serves as input to the next layer.

2.2.1.3.1 Convolutional Neural Networks Convolutional Neural Networks (CNN) [19] are neural networks designed for processing data such as images and video. CNNs excel at identifying features or patterns in such data by using convolution operations, which apply learnable filters over the input data to extract relevant features and local patterns in the data.

2.2.1.4 Hyperparameter Optimization

Hyperparameters are values used to control the behavious of the ML models. For KNN models, the hyperparameter is k, the number of neighbours to conisider before making a decision. For SVM, the hyperparameters include the kernel, the regularization parameter (C) which determines the trade-off between maximizing the marging and accurate seperation of the samples and kernel coefficient (γ) that determines the influence of the support vectors on the shape of the decision boundary. Finding the appropriate hyperparameters are important to ensure the optimal performance of the model. However, these values aren't learned during the training process.

2.2.1.4.1 Grid Search Grid search [20] is a technique used for hyperparameter optimization that uses combinations from predefined list of hyperparameters to determine the optimal combination that maximizes the model's performance. The search algorith generates combinations of predefined values for each hyperparameter of a ML model. Subsequently, the model is trained on each combination, and its performance is assessed over cross-validation. The algorithm then selects the combination that yield the best performance according to the chosen metric for evaluation.

2.2.1.5 Class-Imbalance Problem

When one class in a dataset contains significantly more samples than others, it results in bias in the ML model towards the majority class, leading to poor performance of the models in classifying the minority classes. Class imbalances typically can be addressed using resampling or data augmentation techniques. Two resampling approaches are used in this thesis. The first approach randomly oversamples the minority classes to meet the sample size of the majority class. The second approach randomly omits samples belonging to all the classes to match the sample size of the class with the smallest sample size. However, the resampling methods introduce challenges for generalizing the ML models. ML models could overfit to the redundant samples in the minority classes due to oversampling, and undersampling could lead to information loss as majority class samples are randomly omitted. The challenges posed by the two methods could be overcome using Synthetic Minority Oversampling Technique (SMOTE) [21].

SMOTE helps balance the classes by generating synthetic samples within the neighbourhood of existing minority class samples [22]. In this method, a random minority class sample is selected, and its k-nearest neighbours are found [22]. Then, random values between 0 and 1 are multiplied by the differences between the selected minority class sample and its k neighbours [22]. The products are then added to the selected minority class sample to generate k synthetic samples [22]. The process is repeated until all the classes in the original dataset are balanced [22].

2.2.2 Performance Evaluation of ML Models

The validation method and metrics used to evaluate the accuracy and reliability of the methodologies presented in this thesis are discussed in this section.

2.2.2.1 Cross-Validation

Cross-validation [23] is a popular method for evaluating ML models' performance. This method provides a reliable performance evaluation while helping to mitigate bias. This approach divides the dataset into multiple parts and iteratively trains and tests the model. Each iteration reserves one part for testing the model's performance and the remaining for training the model. Each part is used exactly once to test the model. Finally, the model's average performance is recorded.

2.2.2.2 Accuracy

The accuracy of a ML model is the proportion of accurate classifications the model makes out of all classifications made by the model. True Positives (TP) of a class are the accurate

classifications that belong to the class, while True Negatives (TN) of a class are the accurate classifications that don't belong to the class. Conversely, False Positives (FP) of a class are the incorrect classifications of the class, while False Negatives (FN) of a class are the incorrect classifications that belong to the class. Accordingly, the accuracy of a ML model can be given by:

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

2.2.2.3 Recall

The recall of a class is the proportion of TP out of all predictions of the class. In binary classification, the sensitivity (SN) of the model is the recall of the positive class, and the specificity (SP) is the recall of the negative class. Accordingly, the SN and SP of a binary classification model can be given by:

$$Recall/SN = \frac{TP}{TP + FN}$$

$$SP = \frac{TN}{TN + FP}$$

The recall is important to determine the accurate performance of the model when there is an imbalance among the classes in the dataset. In such cases, the accuracy of the model could be misleading.

2.2.2.4 F1-Score

Precision is the accuracy of the all positive predictions by the model. The precision of a model can be given by:

$$Precision = \frac{TP}{TP + FP}$$

F1-score is another measure that can evaluate the model performance instead of accuracy when there is an imbalance in the number of samples across the classes. It is the harmonic mean of precision and recall; the f1-score is given by:

$$f1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

2.2.2.5 Statistical Significance

Statistical significance is an important metric that helps determine whether the reported performance metrics of ML models are meaningful or simply due to random chance. Statistical significance also helps validate any improvements achieved using the techniques discussed

in this thesis compared to previous works. Hypothesis testing is used to determine the statistical significance where tests such as t-test or Analysis of Variance (ANOVA) determine whether the performance difference between two models is statistically significant.

2.3 Information Measures

In this section, we explore metrics used to measure the relationships between variables in a dataset.

2.3.0.1 Correlation

Correlation is a method used to measure the linear relationship between two variables. It provides the strength and direction of change in one variable with the other. To calculate the correlation between two variables, we divide their covariance, which also measures their relationship, by the product of their standard deviations. While covariance and correlation are measures of linear relationship between the two variables, correlation gives us a standard measure that makes it easier to compare relationships between different data sets. Covariance between X and Y is calculated as the average of the products of the differences between each sample x from X and the mean of X, and its corresponding y from Y and the mean of Y.

$$cov(X,Y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{X}) \cdot (y_i - \bar{Y})$$

$$cor(X,Y) = \frac{cov(X,Y)}{\sigma_X \cdot \sigma_Y}$$

2.3.0.2 Mutual Information

Mutual Information (MI) is another measurement used for the degree of relationship between two variables. Unlike correlation, MI measures both linear and non-linear relationships between the variables [24]. For two discrete variables, MI can be determined using the frequencies of pairs of values appearing in a given dataset [24]. Binning is applied for two continuous variables, where the variables are discretized into discrete bins based on ranges of values, and MI is determined similarly to the discrete variables [24]. An alternative to binning is applying the Nearest Neighbour algorithm, where continuous values are categorized according to the majority in their neighbourhood [24]. Continuous variables are categorized using the Nearest Neighbour approach in the Python library used for implementing the methodologies discussed in this thesis [17].

MI between two variables x and y can be formally interpreted as the reduction in uncertainty of x due to the knowledge of y [25]. A related metric, called Symmetric Uncertainty (SU), provides insight into how both variables contribute to reducing uncertainty for each other, offering a more balanced perspective on their relationship [25]. SU is calculated by normalizing MI with the sum of the entropies of variables x and y [25].

$$MI(X,Y) = H(X) - H(X|Y)$$
$$SU(X,Y) = \frac{2 \cdot MI(X,Y)}{H(X) + H(Y)}$$

In the given equations, H represents the amount of uncertainty, also known as entropy.

2.3.0.3 Cramer's V

Cramer's V (CV) measures the relationship between two categorical variables, calculated using the chi-square statistic [26]. It's derived from another measure called Phi, obtained by dividing the chi-squared value by the sample size [26]. The main distinction is that Phi is applicable only for 2 × 2 contingency tables, whereas CV can be utilized for larger tables [26].

$$\phi = \sqrt{\frac{\chi^2}{N}}$$
$$\phi_c = \sqrt{\frac{\phi^2}{min(k,r) - 1}}$$

Here ϕ denotes Phi, ϕ_c denotes CV, χ^2 denotes the chi-square statistic, N denotes the sample size and k and r represents the number of rows and columns in the dataset. From the equation, we can see that for a 2 × 2 contingency table the CV is equal to Phi.

2.4 Tools and Approach

The methodologies discussed in this thesis are implemented using Python [27]. ML algorithms such as KNN and SVM, hyperparameter optimization techniques such as grid search, and performance evaluation metrics including cross-validation, accuracy, F1-score, SN, SP and recall, and MI implementations are provided by the Scikit-learn library [17]. TensorFlow is used to deploy neural network models [28].

In conclusion, the concepts discussed in this section and the tools establish the framework necessary for achieving our research objectives. By leveraging the concepts and the tools, we propose three ML-based disease diagnosis methods in Sections 3, 4 and 5.

Chapter 3

Alzheimer's Disease Stage Classification using MRI

MRI data is popular in training ML models for Alzheimer's disease diagnosis. Specifically, CNN-based models have shown high accuracy in diagnosing and disease stage classification of Alzheimer's. Besides, studies have also employed brain region segmentation using brain atlases to extract specific regions of the brain known to be associated with Alzheimer's disease to train ML models to achieve high accuracy. However, CNN models require significant computational resources to train, while brain atlases provide a generalized model of the brain, which may not be suitable for Alzheimer's patients due to abnormalities in the brain. At the same time, there are readily available pre-trained CNN-based brain segmentation models that could be used for various applications, including brain region segmentation for Alzheimer's disease diagnosis. This chapter explores ML models' performance on brain region volumes extracted using a pre-trained CNN model and demographic data.

3.1 Introduction 14
3.2 Related Works 15
3.3 Problem Description 16
3.4 Dataset Preparation 16
3.5 Proposed Methodology 17
3.6 Performance Evaluation 19

3.1 Introduction

ML models trained on MRIs are popularly used for Alzheimer's disease diagnosis as they are considered a non-invasive method of diagnosis compared to CSF tests [11]. Popular CNN architectures have achieved accuracy as high as 95.86% in Alzheimer's disease detection and as high as 98.88% in Alzheimer's disease stage classification [6,7]. However, CNNs typically require significant computational resources during the training process. On the other hand, traditional ML models require manual feature extraction methods. While previous works have used brain atlases to extract specific regions of the brain to train traditional ML models, brain atlases are only a generalized representation of the brain regions. Therefore, they might not be suitable for Alzheimer's disease diagnosis as there could be abnormal changes to the brain due to Alzheimer's [12]. This chapter aims to overcome these challenges by using a pre-trained CNN-based brain segmentation model to estimate volumes of brain regions. These estimated volumes are then utilized to train traditional ML models for Alzheimer's disease stage classification.

Clinical studies have shown changes in Alzheimer's patients, such as abnormal atrophy of the hippocampus and expansion of ventricle regions of the brain compared to healthy patients [12]. Previous works have leveraged such abnormalities to train traditional ML models for Alzheimer's disease stage classification. One study trained segmented regions such as the amygdala and hippocampus from MRI scans and achieved 75% accuracy using SVM with RBF kernel [9]. Other works include using Independent Component Analysis (ICA) and probabilistic outputs of SVM and Bayesian classifiers to score segmented regions [10,29]. One such work achieved accuracies as high as 87% for AD vs NL, 78.22% in NL vs MCI and 72.23% in MCI vs AD classifications [29].

In this chapter, we use well-known CNN-based brain segmentation models, such as Brainchop, an open-source web-based volumetric segmentation tool tailored for MRIs [30]. We perform MRI segmentation using the pre-trained weights of a CNN-based model used in Brainchop. Specifically, we estimate the volume of five brain regions associated with Alzheimer's disease. These estimated volumes, along with demographic data like gender and age, are used for training traditional ML models. The remainder of this chapter is organized as follows. Section 3.2 briefly discusses the existing works. Section 3.3 presents the problem statement. Sections 3.4 and 3.5 present the data preprocessing and training methodologies, respectively. Finally, the performance of the models is presented in Section 3.6.

3.2 Related Works

Several previous works have employed CNN models trained on MRI data for diagnosing Alzheimer's disease. Notably, one binary classification approach using CNN achieved 99% accuracy [31]. Chen et al. assessed the performance of 1.5T and 3T MRI for binary classification across various CNN architectures, including LeNet, AlexNet, VGGNet, GoogleNet, and ResNet [6]. The GoogleNet model achieved 95.86% accuracy with 1.5T MRI and 97.15% with 3T MRI for binary classification [6]. In another study, Farooq et al. evaluated the performance of GoogleNet and ResNet in classifying patients into four groups: AD, NL, early MCI (eMCI), and late MCI (lMCI) [7]. The GoogleNet model achieved an accuracy of 98.88% [7]. Adhergal et al. utilized brain atlases, standardized maps of brain regions, to extract the hippocampus region from MRIs and trained them using CNN [8]. Their work achieved 69.53% accuracy for AD vs MCI, 65.62% for AD vs NL, and 66.25% for NL vs MCI classifications [8].

While CNN models are capable of achieving high accuracy, training them typically demands substantial computational resources, often necessitating dedicated hardware such as graphics processing units (GPUs) or tensor processing units (TPUs). Clinical studies have identified changes in specific brain regions, such as the hippocampus and ventricles, due to Alzheimer's disease [12]. For instance, the ventricles of Alzheimer's patients exhibit significant expansion compared to those of healthy individuals over time [12]. Several studies have focused on segmenting brain regions from MRI scans to train traditional ML models, such as SVMs, which generally require less computational power than CNNs. Rallabandi et al. utilized atlas-based methods to segment anatomical brain regions like the amygdala and hippocampus from MRI scans and trained them on decision trees, naive Bayes, KNN, RF, and SVMs with both linear and RBF kernels [9]. The highest accuracy achieved in their work was 75% using an SVM with an RBF kernel [9]. Yang et al. segmented the brain into CSF, white matter (WM), and gray matter (GM) regions and employed ICA to decompose the data [10]. They then used the ICA coefficients to train SVMs, achieving 76.9% accuracy for AD vs NL and 72% for MCI vs NL classifications when trained on the whole brain data, and 80.7% for AD vs NL and 71.1% for MCI vs NL classifications when trained solely on the WM region [10]. Ahmed et al. employed circular harmonic functions (CHF) and bag-of-visual-words (BoVW) techniques to extract the hippocampus region from MRI and represent the extracted content [29]. They trained SVM and Bayesian classifiers on this data, and their probabilistic outputs were further trained on another SVM model for disease stage classification [29]. Their work achieved 87% accuracy in AD vs NL, 78.22% in NL vs MCI, and 72.23% in MCI vs AD classifications [29].

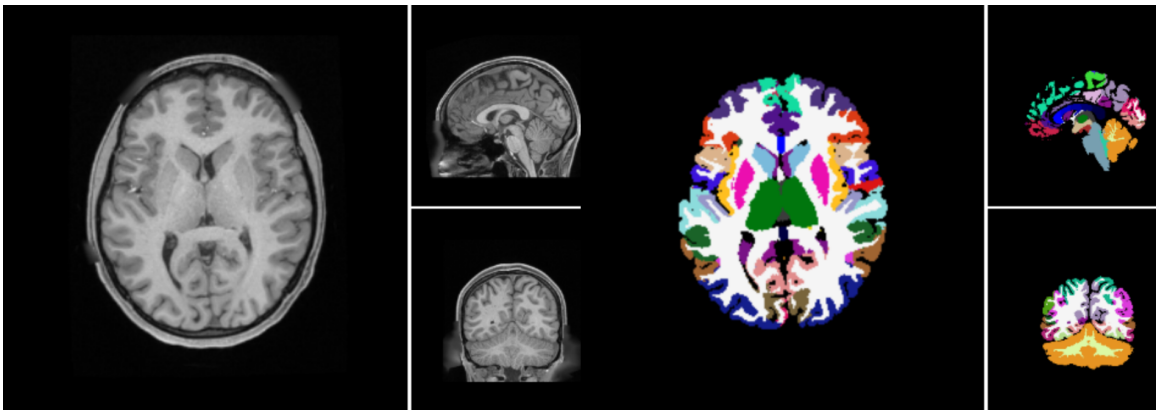

Figure 3.1: MRI before and after segmentation

3.3 Problem Description

While MRI is a non-invasive diagnostic tool compared to CSF tests [11], training CNN models on MRI data often requires substantial computational resources compared to traditional ML models. However, traditional ML models typically rely on manual feature extraction methods. Previous approaches have utilized brain atlases to extract specific brain regions for training ML models such as SVMs. However, brain atlases provide only a generalized representation of brain regions, which might not be accurate for diagnosing Alzheimer's. Alternatively, pre-trained CNN models offer a promising solution for brain segmentation. This chapter explores the performance of ML models trained primarily on volumes of brain regions known to be affected by Alzheimer's disease, estimated through MRI brain segmentation using pre-trained CNN models.

3.4 Dataset Preparation

The MRI scans used in this chapter were obtained from ADNI. Specifically, we use 1.5T MRI scans obtained at baseline when screening ADNI subjects [32]. The dataset consists of 1075 MRI samples, with 525 samples from patients in the MCI stage, 307 from NL, and 243 from patients from AD. Among these samples, 42% belong to female subjects, while 58% belong to male subjects. Before brain region segmentation, the MRI images underwent several preprocessing steps to meet the Brainchop requirements. Initially, the images were normalized to a scale from 0 to 255. Following normalization, they were resized to dimensions of 256x256x256 pixels. Additionally, the voxel size of the images was resampled to 1mm x 1mm x 1mm. The Brainchop preprocessing pipeline consists of further enhancements to the quality of the input volumes by removing noisy voxels and enhancing input volume intensities to improve segmentation accuracy [30].

Table 3.1: Brainchop model Architecture [30]

Layer Type	Output Shape	Parameters	Activation	Details
Input	(256, 256, 256, 1)	0		Input Layer
Conv3D	(256, 256, 256, 21)	588	ReLu	BatchNorm, Dropout
Conv3D	(256, 256, 256, 21)	11,928	ReLu	BatchNorm, Dropout (x7)
Conv3D	(256, 256, 256, 104)	2,288		Output Layer

3.5 Proposed Methodology

The preprocessed MRI data is segmented into 104 brain regions using the pre-trained model obtained from the Brainchop repository. Initially implemented in PyTorch, the Brainchop models were converted to tensorflow.js to facilitate usage in web browsers [30]. The tensorflow.js model and weight files were retrieved from the repository, converted into a single Tensorflow model object, and loaded for segmentation. The inference model implemented in Brainchop is based on MeshNet architecture, consisting of 8 layers [30]. A detailed overview of this architecture is provided in Table 3.1. Additionally, Fig. 3.1 illustrates an example of an MRI segmented using the model; each colour in the output represents a separate brain region. The segmented MRIs will now be used for volume estimation of the regions of interest within the brain.

The hippocampal, parahippocampal, ventricles, entorhinal, and cerebral white matter regions were identified as regions of interest based on clinical studies [12]. Binary maps for each region were obtained from the segmented MRIs. The volumes of these regions were estimated in mm^3 by summing their respective binary map. Demographic data such as the gender and age of each patient were also collected and stored in a tabular format along with the volumes. A random 80-20 split was used to generate train and test datasets. The test dataset was used to train SVM [16] and KNN [15] models. Grid search [20] was employed to identify the best hyperparameters for the models. The SVM model used RBF kernel, and the search included values 0.1, 1 and 10 for C and 0.01, 0.1, 1, $\frac{1}{\text{panel\_size}*\text{variance}}$, and $\frac{1}{\text{panel\_size}}$ for γ. The KNN model's hyperparameters were values of k ranging from 5 to 20. The selection criterion for the grid search was based on model accuracy, with the hyperparameters yielding the maximum accuracy chosen for each model. The models' average accuracy and recalls over 10-fold cross-validation [23] using the train set and the best hyperparameters were recorded. Finally, the models' accuracy and recalls on the test set were recorded to assess their performance. Fig. 3.2 visually represents the framework discussed in this chapter.

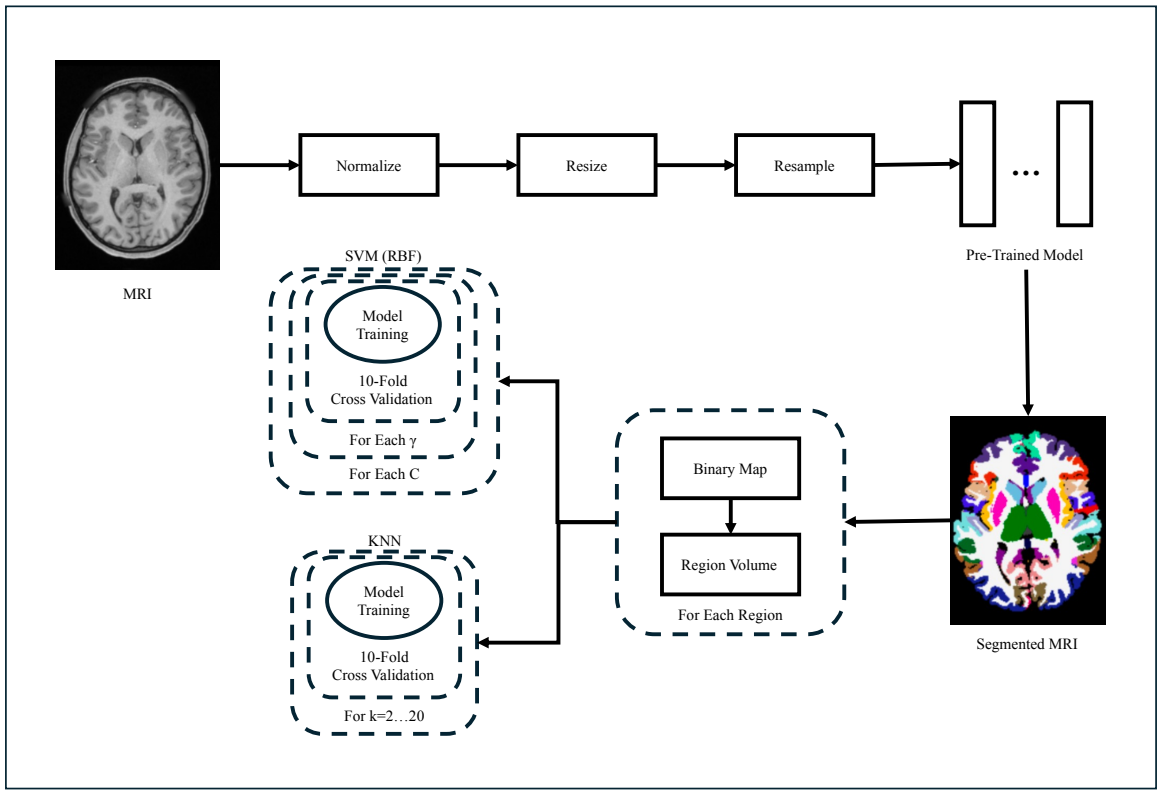

Figure 3.2: Overall Framework for Disease Stage Classification using MRI.

Table 3.2: Comparison of Performances of MRI-based Methods [7–10, 29]

Work	**Approach**	**Accuracy**
Rallabandi et al.	Cortial Thickness of Amygdala and Hippocampus	75%
Adhergal et al.	Hippocampus	67.13% (average)
Yang et al.	White Matter + CSF + Gray Matter	74.45% (average)
Yang et al.	White Matter	75.9% (average)
Ahmed et al.	Hippocampus	79.15% (average)
Farooq et al.	GoogleNet	98.88%
This Chapter	**Volumes of 5 Brain Regions + Demography**	**99.74%**

3.6 Performance Evaluation

During the validation and testing, the SVM model achieved remarkable performance, with an accuracy of 99.74%. Notably, it achieved 100% recall for AD and NL samples and 99.47% for MCI samples. The best hyperparameters for the SVM model are C=10 and γ=0.01. In contrast, the KNN model showed lower accuracy, achieving only 54.46% during validation, with 12.66%, 32.31%, and 86.24% recalls for AD, NL, and MCI samples, respectively, using k=18. During testing, the KNN model achieved a lower performance. The training and testing processes were repeated after implementing random oversampling to address class imbalance in the dataset. Following this, the SVM model demonstrated 100% accuracy during both validation and testing. The KNN model improved significantly, achieving 100% accuracy during validation and 97.51% during testing. It achieved 100% recall for AD and NL samples and 92.53% for MCI samples. Table 3.2 presents the best performance reported in this chapter, alongside the performances from previous works.

As discussed in section 2.2.2.5, we conducted hypothesis testing using a t-test to evaluate the statistical significance of the best performance reported in this chapter. Specifically, we compared the recalls and the accuracy of the SVM model with those of the GoogleNet model discussed in [7]. The test indicated that the difference in performances is statistically significant. While the reported performance metrics show that our model outperforms the previous works, the significance test verifies that the reported metrics are unlikely due to random chance. Therefore, we move to the next chapter, assured by the strong and validated performance achieved using the first approach of this thesis.

Chapter 4

Identifying Biomarker Panels for Early Detection of Alzheimer's Disease using Blood

While the MRI-based method discussed in Chapter 3 is highly accurate in diagnosing Alzheimer's disease, it remains expensive compared to other imaging solutions such as PET. Moreover, it might not be practical for patients requiring medical devices such as pacemakers due to MRI's interaction with metallic foreign objects. This chapter introduces two novel feature selection methods to identify relevant blood features for identifying early cases of Alzheimer's. In the first method, we rank features based on their dependency on the diagnosis using metrics such as MI, SU, and CV. Panels are formed by iteratively selecting top features and gradually increasing the panel size. We then propose a second method that determines the relevant features by estimating the Euclidean distance between the class means for each feature, employing a threshold to filter out irrelevant features. Candidate panels identified using each method are extensively tested on two datasets, and their performances are reported.

4.1 Introduction 21
4.2 Related Work 23
4.3 Problem Description 24
4.4 Dataset Preparation 25
4.5 Proposed Feature Selection Methods 25
4.5.1 Dependency-Dependent Feature Selection Method 25
4.5.2 Euclidean Distance-based Feature Selection Method 27

4.6 Performance Evaluation 30

4.1 Introduction

While the MRI-based approach discussed in Chapter 3 is highly accurate in Alzheimer's disease diagnosis, it remains expensive when compared to beta-amyloid imaging using PET. Furthermore, patients need thorough foreign body screening before undergoing an MRI as its interaction with metallic foreign objects poses a risk [33]. Therefore, patients relying on medical devices such as pacemakers cannot undergo MRI for diagnosis [33]. On the other hand, ML models trained on blood proteomic data have demonstrated high accuracy in identifying early cases of Alzheimer's [11]. A blood-based diagnostic tool is cost-effective compared to imaging procedures such as MRI or PET but, at the same time, remains non-invasive compared to the CSF test [11]. However, reproducing results using panels identified in previous studies has proven challenging due to inconsistent performance. One contributing factor to this challenge is the inherent class imbalance problem with the dataset used, and studies remain opaque regarding the solutions used to address this problem. This chapter introduces two novel feature selection approaches and discusses how these methods, paired with well-known data augmentation techniques, can mitigate the challenges associated with Alzheimer's diagnosis using blood.

Several previous works have used blood proteomic data to identify relevant biomarker panels that could help in identifying early cases of Alzheimer's. Merit is one such approach where panels are scored such that the panels are rewarded for a high correlation between the features in the panel and the diagnosis [11]. At the same time, the panels are penalized for high correlation among the features in the panel [11]. In this work, the models were trained on AD and NL samples obtained at baseline and tested on MCI and NL samples obtained 12 months from baseline [11]. The best-performing model achieved an average of 81% and 71.5% SN and SP, respectively [11]. In comparison, Alzheimer's disease can be diagnosed by a human expert with 81% and 70% SN and SP, respectively [11]. Another approach involves forming panels of all possible combinations of relevant features identified using a statistical t-test [34]. This approach restricted The panel sizes to 2, 3, 4 and 5 [34]. A panel of 5 features achieved the highest performance of 86.5% and 82.1% SN and SP, respectively [34]. All panels formed that met minimum performance criteria were further tested rigorously for 1000 iterations to test their reliability [34]. The 5-feature panel achieved 77% in this test [34].

Other approaches to finding relevant features include using classification accuracy of individual features, univariate and multivariate analysis of features and correlation [35–37]. Area Under the Curve (AUC) of ML models trained on individual features is used to

Table 4.1: Optimal Panels Identified in [11]

Panel	Train SN	Train SP	Test SN	Test SP
A2M, ApoE, BNP, Eot3, PLGF, RAGE, SGOT	88.5	70.4	80.1	70.4
A2M, ApoE, BNP, Eot3, PYY, RAGE, SGOT	88.9	73.8	77.9	74.1
A2M, ApoE, Eot3, IGM, MCSF1, PYY, RAGE, SGOT	85.3	71.6	83.8	70.4
ApoA2, ApoE, BNP, BTC, Eot3, PYY, RAGE, SGOT	85.0	75.0	80.1	72.2
A1M, A2M, ApoE, BNP, BTC, Eot3, IGM, MCSF1, PAPPA, SGOT	88.1	72.9	83.1	70.4

identify relevant features in one work where all possible combinations of features with an AUC greater than 0.6 are used to train the models [35]. In this work, a six-feature panel achieved the best performance of 85.4% and 78.6% SN and SP, respectively [35]. Another work employed univariate analysis using Analysis of Covariance (ANCOVA) to determine individual features relevant to disease diagnosis and multivariate analysis to identify the best feature panel for disease diagnosis [36]. While the five individual features identified in the study yielded 50% SN, a 14-feature panel identified during the multivariate analysis achieved 86.5% and 84.2% SN and SP, respectively [36]. Finally, a study that trained 11 statistically significant features identified using Pearson correlation and t-test and demographic data on the RF model achieved 74% and 91% SN and SP, respectively [37].

We emphasize the importance of appropriate feature selection methods to improve the model training efficiency of Alzheimer's disease detection and introduce two feature selection methods to identify robust and reproducible feature panels in this section. The first method used metrics such as MI, SU and CV to estimate the relationship between features and the diagnosis and rank the features accordingly. Panels are generated in this method by iteratively increasing the panel size and adding the top features. Conversely, the second method aims to identify relevant features by estimating each feature's Euclidean distance between the class means. The remainder of this chapter is organized as follows. Section 4.2 discusses the existing research works briefly. Section 4.3 introduces the problem description. Section 4.4 and Section 4.5 discuss the dataset preparation and the feature selection methods, respectively. Finally, the performance of the proposed feature selection methods is evaluated in section 4.6.

4.2 Related Work

The number of works identifying relevant blood protein panels to accurately train ML models on blood data has been on the rise as a non-invasive and inexpensive alternative to imaging and CSF tests to diagnose Alzheimer's [11]. One such approach uses Merit, a heuristic that rewards for a high correlation between features in a panel and the diagnosis and penalizes for a high correlation among the features, where candidate panels are formed through brute force using feature subsets pre-selected using the Merit [11]. Here, SVM models using RBF are trained using the panels on AD and NL samples obtained at a baseline and are validated using 10-fold cross-validation [11]. Models are filtered out using minimum performance criteria of 81% SN and 80% SP [11]. Furthermore, the models are tested on MCI and NL samples obtained 12 months from the baseline [11]. The best-performing panels identified in this study are summarized in Table 4.1. Another work by Eke, Jammeh, et al. emphasizes reproducing the panel performances and recommends reliability testing [34]. All possible combinations of 14 statistically significant features determined through t-tests were used to form panels of 2, 3, 4 and 5 features [34]. The panels were trained on SVM using RBF kernel and evaluated using 10-fold cross-validation [34]. Panels that didn't meet minimum performance criteria were discarded, and the remaining candidate panels were rigorously tested for 1000 iterations to determine their reliability [34]. In each iteration, 10-fold cross-validation was carried out, and the number of iterations where the model met the minimum performance criteria was used to determine the reliability [34]. A panel of five features had the best performance in this work, achieving an average of 86.5% SN, 82.1% SP and 77.8% reliability over the 1000 iterations [34].

A work by Jammeh, Zhao et al. aims to develop a revolutionary point-of-care tool for Alzheimer's diagnosis using blood samples [35]. The work employs classification accuracy of individual features to identify the best panel to aid the diagnosis [35]. Each feature is initially trained on Naive Bayes, Logistic Regression, Multivariate Perceptron, RF and SVM models [35]. All possible combinations of features that achieved an AUC over 0.6 were generated using brute force [35]. The panels were trained on the five learning models, their performance was assessed, and relevant panels were identified using a minimum performance criterion of 75% SN and SP [35]. Llano et al. employs univariate and multivariate analysis to identify relevant feature panels for diagnosis and prognosis of Alzheimer's [36]. ANCOVA is used to identify relevant features that could differentiate different disease stages, while a separate multivariate analysis is used to identify relevant feature panels [36]. Relevant features are filtered out using Student's t-test, and panels are formed using relative importance and the maximum predictive accuracy of features on RF, partial least squares and bagging during internal validation using two-thirds of the original dataset [36]. A limit of 80% was applied to the pair-wise correlation among the features in the panels during panel forma-

tion [36]. Candidate panels were then tested on RF, SVM with RBF kernel, Neural Network, Partial Least Squares, bagging and KNN models, and their performance was assessed [36]. O'Bryant et al. explore feature selection using the statistical significance of features using blood data from two studies [37]. The work identifies 11 statistically significant features with a high correlation with the diagnosis [37]. The relevant features were trained on the RF model using the Texas Alzheimer's Research and Care Consortium (TARCC) dataset and validated using ADNI data [37]. While the model initially achieved 54% SN and 78% SP, the performance was improved to 74% SN and 91% SP after incorporating demographic data [37]. A model trained on CSF achieved 84% SN and 100% SP in comparison [37].

Using data from different sources during training and validation may be counterproductive due to potential differences in data collection methods, population distribution, and demographics. Such differences could introduce biases and affect the performance of the model. Hence, this chapter solely relies on data from ADNI. The dataset comprises 190 proteins extracted from EDTA plasma samples obtained from ADNI patients; these proteins are considered to be associated with various diseases and cellular signalling pathways [38]. Over one thousand sixty-five patients underwent overnight fasting, and their blood samples were collected the next morning [38]. Through centrifugation, EDTA plasma was isolated from anti-coagulated blood samples to prevent clotting [38]. These samples were then shipped to Rule-based Machine for analysis [38]. The dataset consists of 396 MCI, 112 AD, and 58 NL samples collected at baseline and one-year follow-up [38]. More than half of the samples are male participants, with the average baseline age being 75 [38].

4.3 Problem Description

In Chapter 3, we showed that models trained on volumes of brain regions extracted from MRI can be an accurate and non-invasive tool for diagnosing Alzheimer's. However, it poses risks for patients with medical devices like pacemakers due to interactions with foreign metal objects [33]. Moreover, MRI is expensive compared to beta-amyloid imaging using PET and CSF test. ML models have shown promising results in early diagnosis using blood proteomic data, a relatively non-invasive and cost-effective method [11]. However, the performance achieved with prior works' biomarker panels is often not reproducible. Moreover, there's a need for efficient feature selection methods to identify such biomarker panels and developing a blood-based method could also pave the way for understanding disease pathology [11]. This chapter aims to introduce feature selection techniques to identify robust and reproducible biomarker panels that effectively differentiate between AD and NL patients.

4.4 Dataset Preparation

Forty-four incomplete features out of 190 in the dataset were removed. Initially, outliers in the data were removed using the z-score approach. However, this led to more than half of the samples being removed. Instead, the data were standardized using z-score. Two sets of data were initially generated. The first consists of a training dataset of AD and NL samples obtained 12 months from baseline and a test dataset of MCI and NL samples obtained at the baseline. The rationale behind this set is to assess the ML models' performance in diagnosing early cases when trained exclusively on control and dementia patients. The second set consists of training and test datasets of samples from all three classes, irrespective of when the samples were obtained. Both MCI and AD samples in this set are treated as a single class under the assumption that while MCI patients don't have dementia yet, they still exhibit characteristics indicative of Alzheimer's disease. The second set simulates the data encountered in practical scenarios where samples belonging to all three classes would exist simultaneously. One major problem while training the ML models is the bias towards MCI during classification. The primary factor contributing to this is the class imbalance within the ADNI dataset. Specifically, the dataset has 112 NL, 209 AD and a much larger 742 MCI samples. This chapter addresses class imbalance using data augmentation methods such as Random oversampling, undersampling and SMOTE [21]. Following the augmentation of the two initial sets of data, we now apply feature selection across all 12 datasets.

4.5 Proposed Feature Selection Methods

This section proposes two feature selection methods. The first method proposed is the dependency-dependent feature selection method. Then, the section introduces our feature selection method based on Euclidean distance.

4.5.1 Dependency-Dependent Feature Selection Method

This chapter's first feature selection method ranks the features according to their relationship with the labels, measured using MI, SU and CV. Traditional methods, such as Pearson correlation, might not be suitable for the dataset as they measure the linear relationship between two continuous variables. In comparison, MI estimates the measure of dependence between two variables irrespective of their type. SU is an extension of MI, where the measure of dependence is normalized. CV measures the dependency between two discrete variables.

First, the six datasets are discretized using a multi-interval discretization algorithm to measure CV. The algorithm is an extension of the binary discretization algorithm discussed in [39]. In the binary discretization algorithm, the features are categorized using cutpoints,

a threshold value, where values greater than the cutpoint are grouped as one category and the rest as the other [39]. The optimal value for the cutpoint from a list of candidate values is determined using MI, where categorizing a feature using the optimal value would yield the highest MI between the feature and the label [39]. The multi-interval classification approach used in the chapter allows multiple categories using more than one cutpoint equidistant from each other. The optimal number of categories is determined in the modified approach rather than the specific values for each cutpoint. The optimal number of categories is determined for each feature by iteratively increasing the number of categories from 2 to n and generating n-1 equidistant cutpoints, categorizing the values accordingly and measuring the MI in each iteration. The optimal number of categories yields the highest MI here. Algorithm 4.1 shows the multi-interval discretization method.

Algorithm 4.1: Multi-Interval Discretization

Input: n, *labels*, *features*
Output: *features*
foreach *feature in features* **do**
  $best\_mi \leftarrow -\infty$;
  $best\_feature \leftarrow$ None;
  for $num\_intervals$ *in* $2..n$ **do**
    $cut\_points \leftarrow \text{linspace}(\min(feature), \max(feature), num\_intervals)$;
    $discrete\_feature \leftarrow \text{categorize}(feature, cut\_points)$;
    $mi \leftarrow \text{mutual\_information}(discrete\_feature, labels)$;
    if $mi > best\_mi$ **then**
      $best\_mi \leftarrow mi$;
      $best\_feature \leftarrow discrete\_feature$;
    end
  end
  $feature \leftarrow best\_feature$;
end

The features are ranked according to their degree of relationship with the labels, measured using MI, SU and CV for all 12 datasets. Panels are formed iteratively by increasing the panel size, using top n features, where n is the panel size. In each iteration, KNN [15] and SVM [16] (with a polynomial kernel of degree 2) models are trained using the panel and their average SN, SP and accuracy over 10-fold cross-validation [23] are recorded. An exit criterion of a 2% improvement in SN, SP or accuracy over three iterations is applied. If none of these metrics improve by at least 2% over three iterations, then n - 3 will be considered the optimal number of features, and the loop will be terminated. This exit criterion is similar to the validation-based early-stopping technique commonly used in training neural networks [40].

Candidate panels are identified by filtering out the panels formed that don't meet a

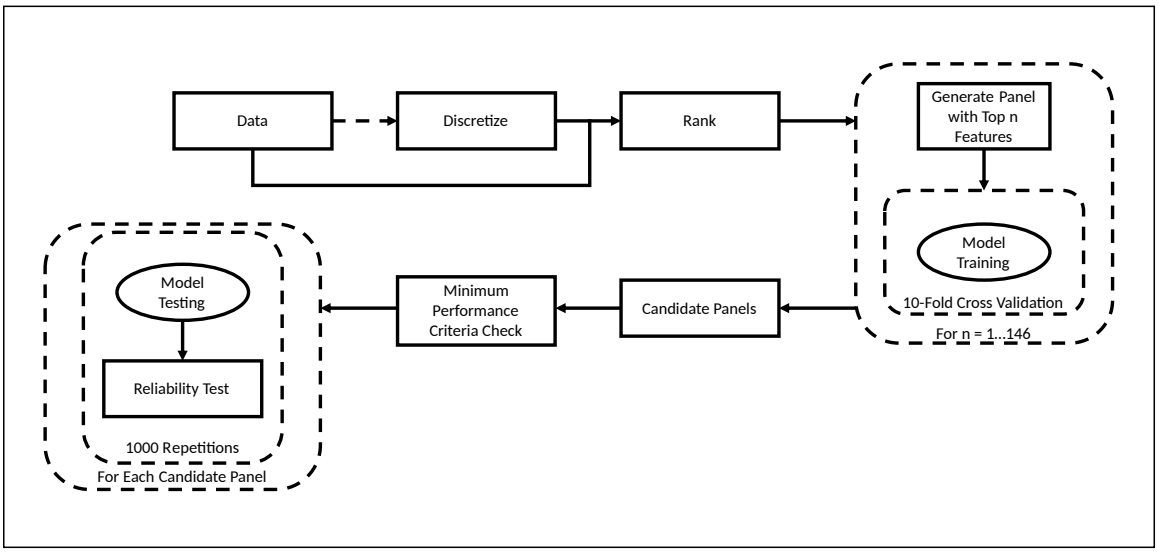

Figure 4.1: Overall Framework for Dependency-Dependent Feature Selection.

minimum performance criteria of 75% SN and SP. The candidate panels then undergo rigorous testing, where their reliability is measured. The testing phase is similar to the reliability test discussed in [34], where KNN [15] and SVM [16] models are trained on each candidate panel for 1000 iterations. The models' average SN, SP and accuracy over 1000 iterations are recorded. Reliability is the percentage of times the models meet the minimum performance criteria throughout the 1000 iterations [34]. Fig. 4.1 illustrates the overall framework for this feature selection method.

4.5.2 Euclidean Distance-based Feature Selection Method

This chapter's second method uses the Euclidean distance between class means for each feature to identify relevant features. The underlying concept is that the samples would cluster around their respective means within each class, resulting in a shorter distance between the means of a class and a sample within that class compared to the distance between the means an a sample from another class. As a result, for a relevant feature for distinguishing between classes, the means of each class should have a noticeable difference. However, detecting this pattern becomes challenging in higher dimensions where distances tend to become uniform. Therefore, this method evaluates the class means of each feature separately.

To further explain this method, consider Fig. 4.2. The figure shows a feature space containing features f_1, f_2 and f_3 from various perspectives, with sample classes differentiated by colours. While there is no apparent distinction among the samples in the three-dimensional space, two distinct clusters emerge when observed in two dimensions. Additionally, upon examining the f_1f_2-plane and the f_1f_3-plane, it becomes apparent that samples from both

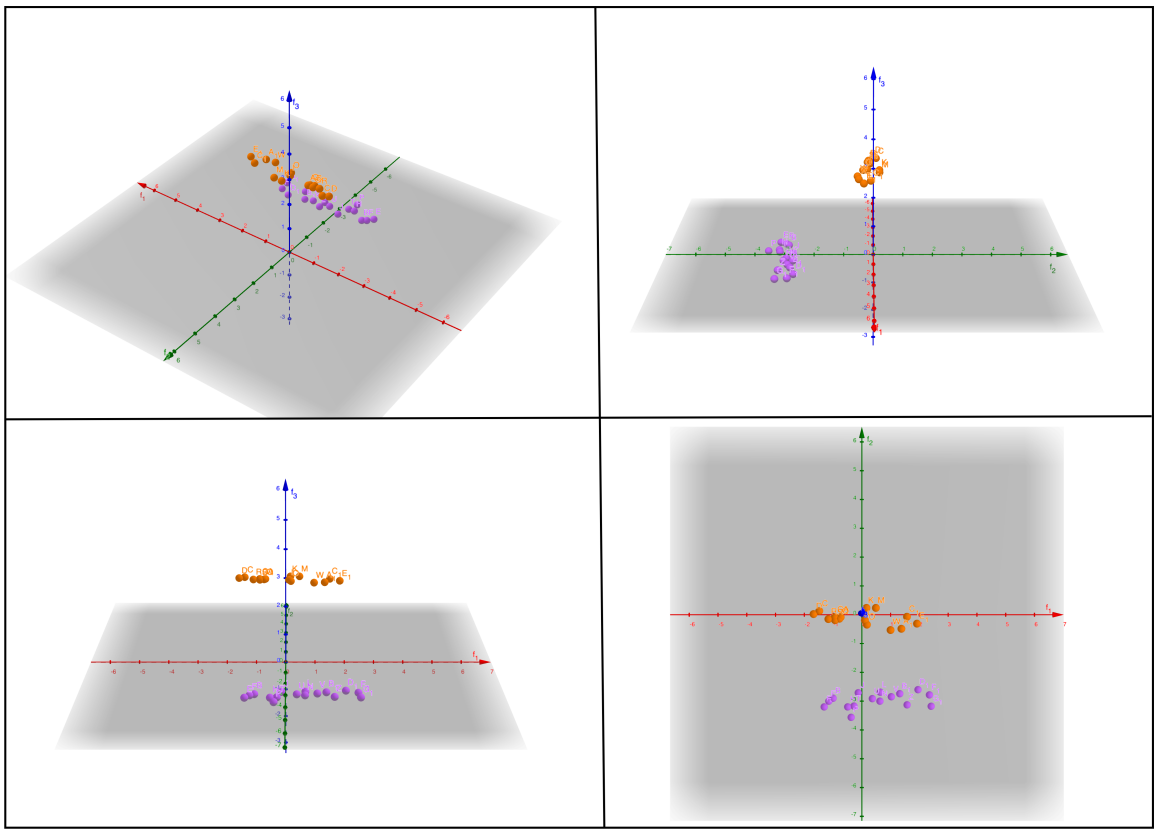

Figure 4.2: Feature space, f_3f_2-plane, f_1f_2-plane and f_1f_3-plane (clockwise).

classes overlap in f_1. As a result, the class means in f_1 would be similar for all classes while noticeably different in f_2 and f_3. Using the Euclidean distance-based feature selection approach, f_1 would be removed. Fig. 4.3 illustrates that the class means are noticeably different for some features in the dataset used in this chapter. Meanwhile, the overlap in class means illustrated in Fig. 4.3 suggests that the features may be irrelevant for classification.

Panels are formed by filtering out irrelevant features by applying a minimum threshold value on the Euclidean distance between the class means for each feature. The samples are separated according to their respective classes, and class means and distance for each feature are calculated. The threshold values are iteratively increased from the minimum to the maximum distances among the features in ten equal iterations. In each iteration, the relevant features, determined by the threshold, are used to train KNN [15] and SVM [16] models, and their average SN, SP and accuracy over 10-fold cross-validation [23] are recorded. A 2% exit criterion is also applied in this method. Candidate panels are identified by filtering out panels that fail to meet the minimum performance criteria and undergo testing following the methodology discussed in the Dependency-dependent feature selection approach. Fig. 4.4 illustrates the overall framework for this feature selection method.

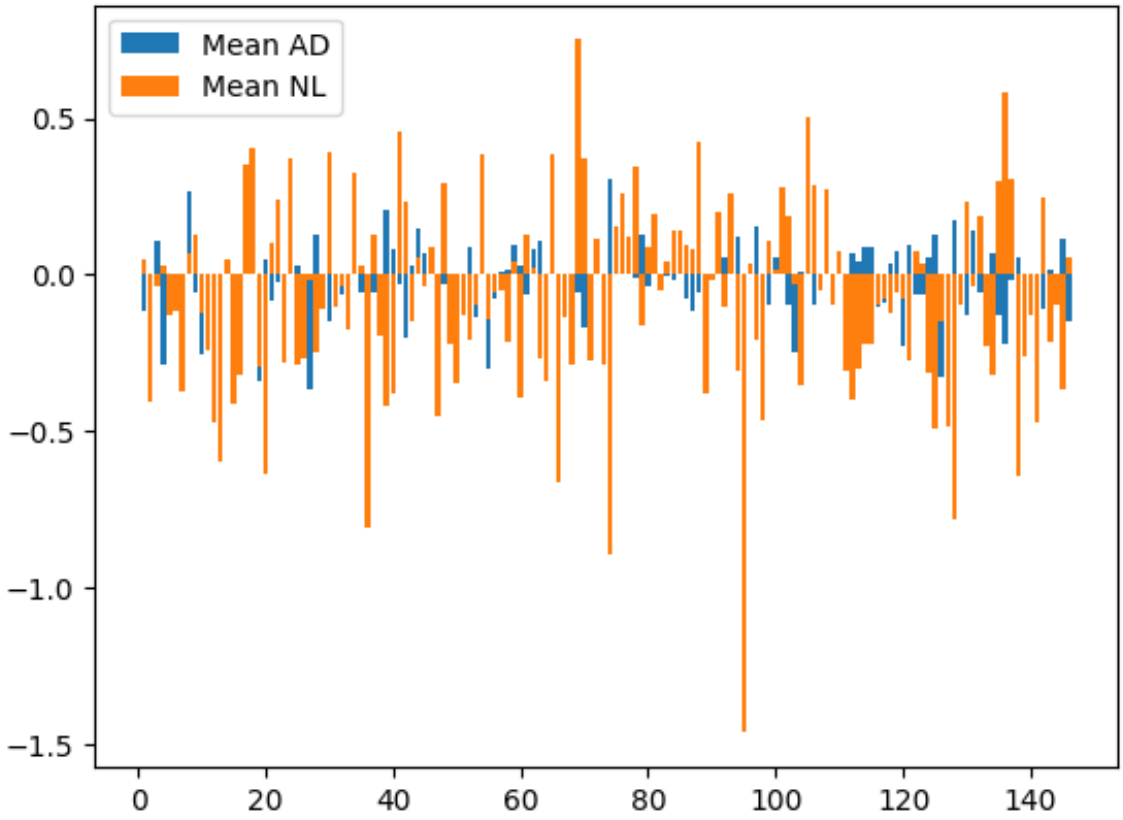

Figure 4.3: Class means for each feature.

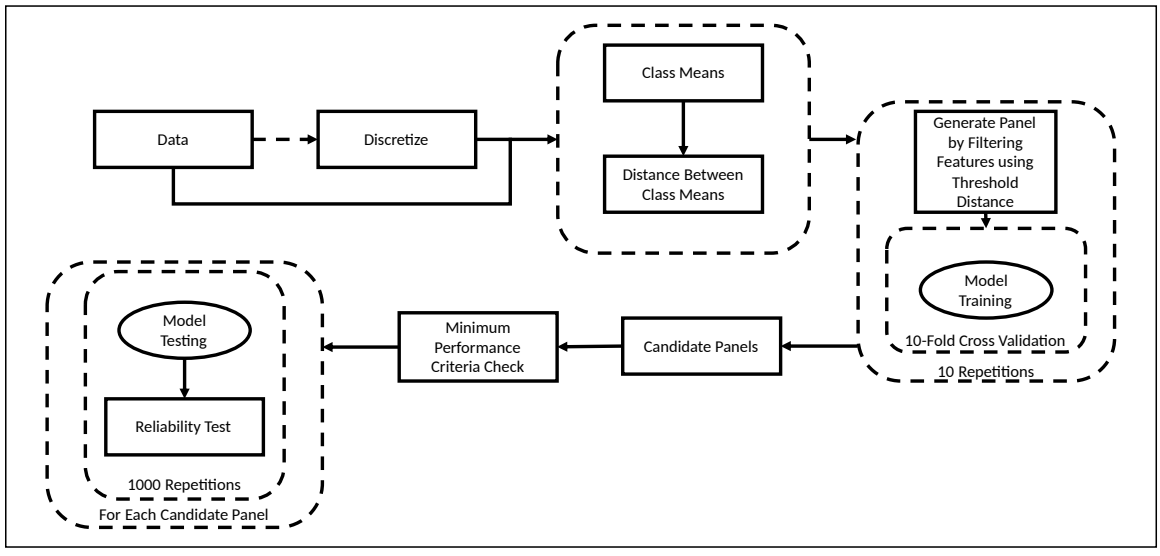

Figure 4.4: Overall Framework for Euclidean Distance-based Feature Selection.

4.6 Performance Evaluation

We used the panels identified in [11] to train SVM models on three augmented datasets derived from set 1, aiming to establish a solid foundation for comparison with the performance of our models. The models consistently displayed a bias towards the AD class, achieving higher SN than SP. Performance metrics varied across datasets, with the model trained on the oversampled dataset achieving 79% SN and 52% SP, the undersampled dataset achieving 60% SN and 68% SP, and the SMOTE dataset achieving the best performance of 70% SN and 79% SP. Set 1 was specifically selected to align with the experimental conditions of the previous work.

We identified 340 candidate panels that meet the minimum performance criteria through the dependency-dependent feature selection method. Among them, 210 panels were identified using set 1, while 130 were identified using set 2. Unfortunately, none of the panels validated on set 1 met the minimum performance criteria during testing. During validation, the best-performing panel for set 1 contained 22 features, used random oversampling for data augmentation and was trained on an SVM model with a polynomial kernel of degree 2. The features in this panel were ranked using SU. The panel achieved 95.84% accuracy with 98.12% SN and 92.43% SP during validation. However, during testing, the model achieved only 80.62% accuracy with 91.18% SN and a noticeably lower 13.79% SP. Fig. 4.5 depicts the SN, SP and overall accuracy during testing for this feature selection method using set 1. The best-performing panel for set 2 achieved 89.67% accuracy with 90.16% SN and 85% SP during testing. This panel contains 20 features, uses random oversampling for data augmentation, and was trained on an SVM model with a polynomial kernel of degree 2. The features in the panel were discretized before being ranked using SU. Fig. 4.6 depicts the SN, SP and overall accuracy during testing for this feature selection method using set 2.

We identified 209 candidate panels that meet the minimum performance criteria through the Euclidean distance-based feature selection method, with 123 panels identified using set 1 and 86 using set 2. Similar to the panels identified using the dependency-dependent method, none identified using set 1 met the minimum performance criteria during testing. During testing, the top-performing panel for set 1 achieved 88.33% accuracy with 98.74% SN but a notably low SP of 17.24%. This panel consisted of 130 features, employed random oversampling for data augmentation and was trained on SVM with a polynomial kernel of degree 2. During validation, the best-performing panel for set 1 consisted of 71 features and achieved 98.87% accuracy, 98.74% SP and 94.26% SN when trained on an SVM with a linear kernel. However, the accuracy dropped to 79.74% during testing, with 89.9% SN and a significantly lower 10.34% SP. The features in the panel were discretized, and random oversampling was used for data augmentation. Fig. 4.7 shows the SN, SP and overall

accuracy during testing for this feature selection method using set 1.

For set 2, the highest-performing panel achieved 99.53% accuracy during testing, with 100% SN and 95% SP. This panel comprised 133 features. A panel of 65 features achieved 99.06% accuracy with 99.48% SN and 95% SP, and another panel of 48 features achieved 98.56% accuracy. The panels were augmented using random oversampling and were trained on SVM with a polynomial kernel of degree 2. Fig. 4.8 shows the SN, SP and overall accuracy during testing for this feature selection method using set 2. The accuracy trent with decreasing threshold for this dataset during testing is shown in Fig. 4.9. The figure shows an initial rapid increase in accuracy that stabilized at 0.15, with only a marginal increment for subsequent lower threshold values. Table 4.2 presents the best performance reported in this chapter, alongside the performances from previous works.

The statistical significance of the best-performing panels during testing reported in this chapter was validated using the performance of an 8-feature panel from [11] through a t-test. Specifically, we validated the performance of the 20-feature panel identified using the dependency-dependent method and the 65-feature panel identified using the Euclidean distance-based method on set 2. While the performance difference between the Euclidean-distance-based method and the previous work showed statistical significance according to the test, the performance difference between the dependency-dependent method and the previous work didn't. This lack of significance may be due to the methodological and performance similarities between the dependency-dependent method and Merit used in [11]. Merit uses SU to score features, similar to how SU is used in ranking the 20-feature panel [11]. Moreover, the performance reported for the dependency-dependent method is similar to that reported in [11]. At the same time, the Euclidean distance-based feature selection method emerges as a novel approach for identifying optimal feature panels for disease diagnosis. The next chapter broadens the scope of this chapter by discussing the application of the two feature selection methods in disease stage classification.

Table 4.2: Comparison of Performances in Alzheimer's Diagnosis using Blood [11, 34–37]

Work	**Approach**	**Panel Size**	**SN**	**SP**
Eke, Jammeh, et al.	Merit	8	83.8%	70.4%
Eke, Jammeh, et al.	t-test	5	86.5%	82.1%
Jammeh, Zhao, et al.	AUC	6	85.4%	78.6%
Llano et al.	Uni/Multivariate Analysis	14	86.5%	84.2%
O'Bryant et al.	t-test, Pearson Correlation	12	74.0%	91.0%
This Chapter	**Previous Panel**	**8**	**70%**	**79%**
This Chapter	**Dependency-Dependent**	**20**	**90.16%**	**85%**
This Chapter	**Euclidean Distance**	**65**	**99.48%**	**95%**

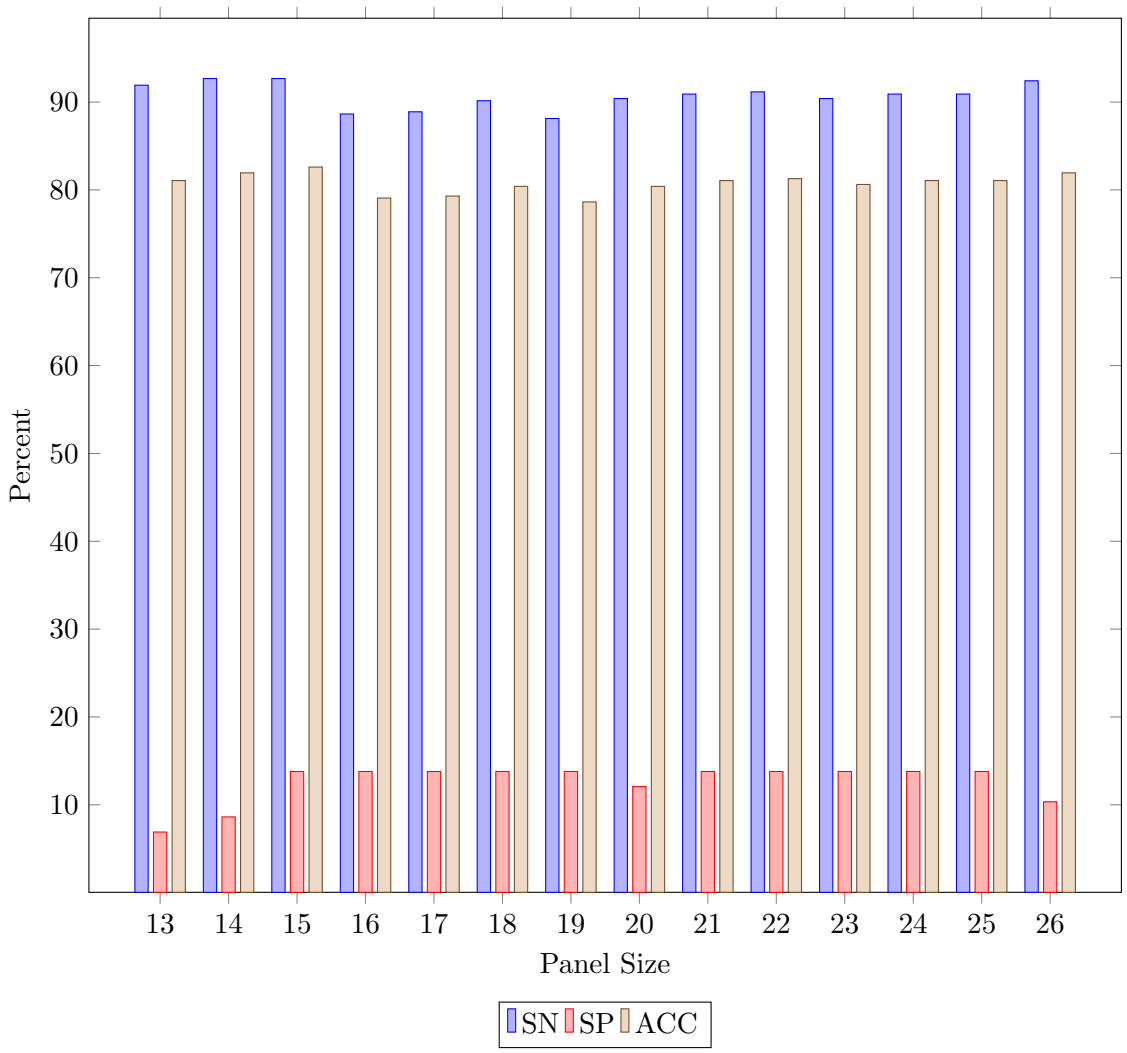

Figure 4.5: SN (sensitivity), SP (specificity), and accuracy (ACC) for Set 1 during testing (dependency-dependent method).

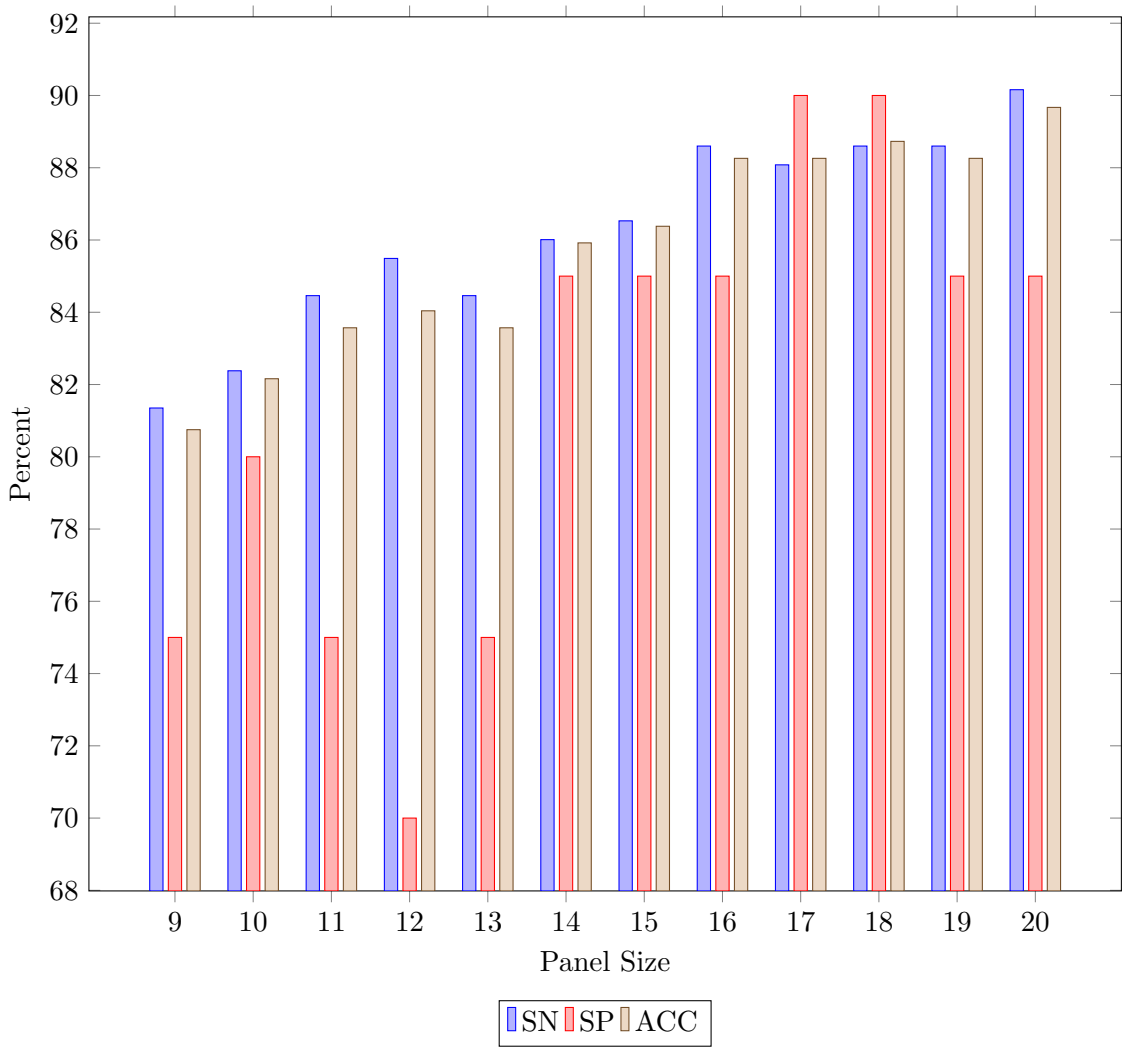

Figure 4.6: SN, SP, and accuracy for Set 2 during testing (dependency-dependent method).

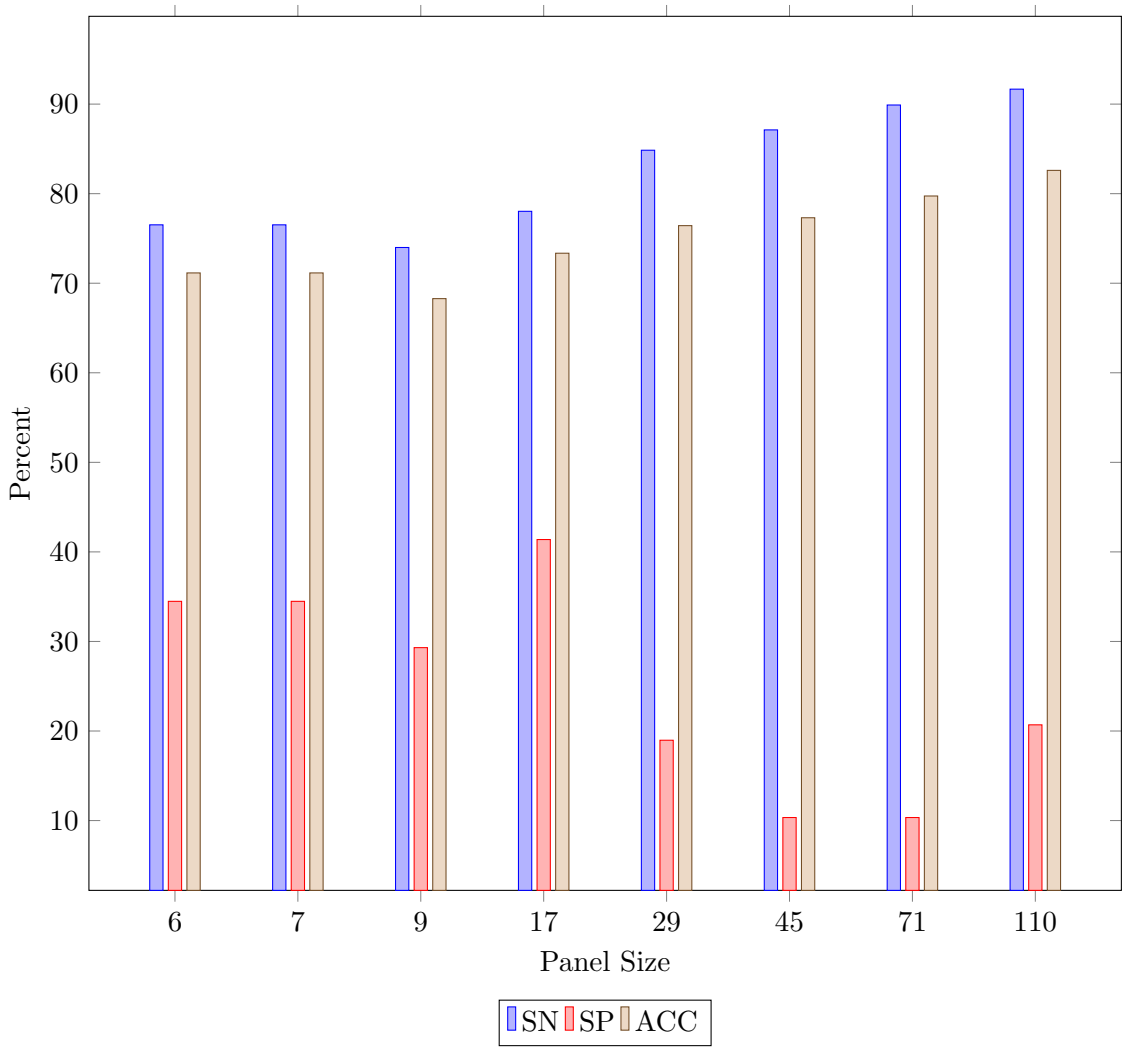

Figure 4.7: SN, SP, and accuracy for Set 1 during testing (Euclidean-Distance method).

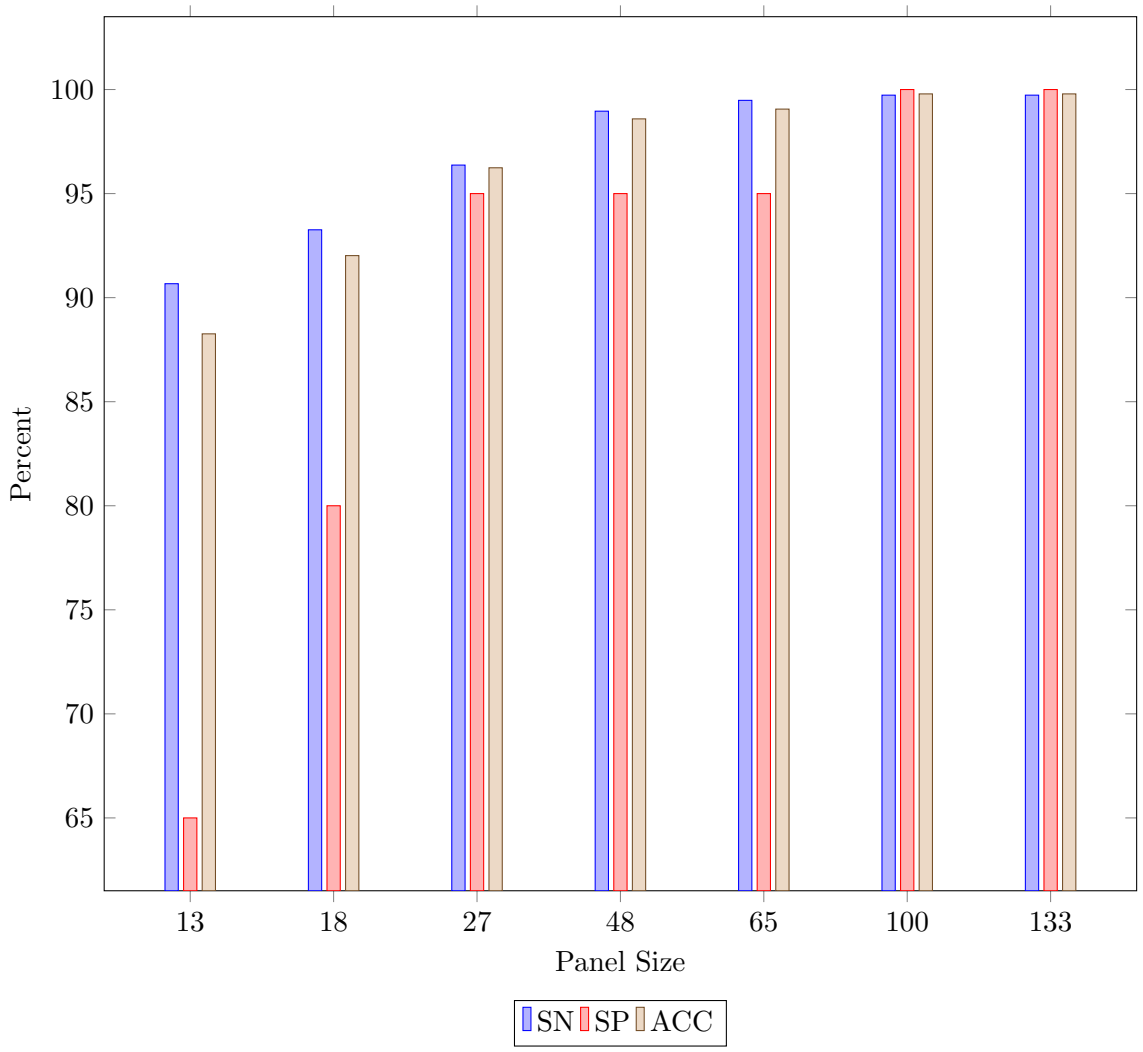

Figure 4.8: SN, SP, and accuracy for set 2 during testing (Euclidean-Distance method).

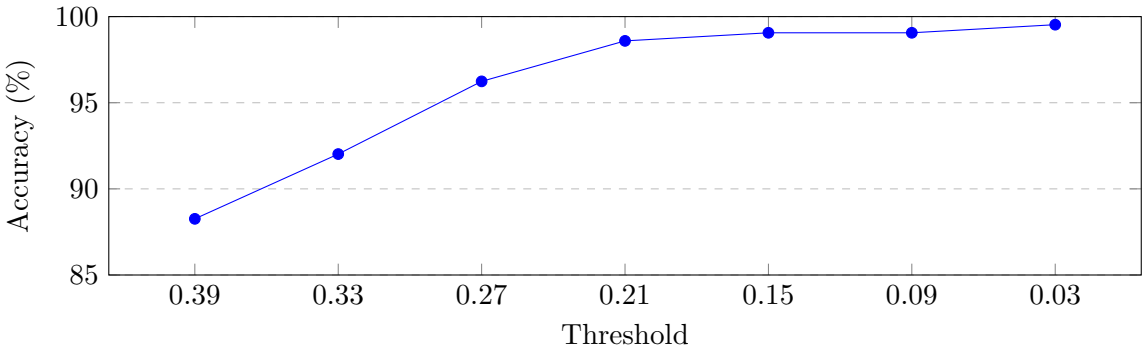

Figure 4.9: Testing Accuracy for different threshold values.

Chapter 5

Identifying Biomarker Panels for Alzheimer's Disease Stage Classification using Blood

While imaging-based approaches, such as the method discussed in Chapter 3, are highly accurate in the disease stage classification of Alzheimer's, blood-based approaches, such as the method discussed in Chapter 4, primarily focus on distinguishing patients with Alzheimer's from normal individuals rather than classifying them into their specific disease stage. In this chapter, we evaluate the performance of ML models in the disease stage classification of Alzheimer's patients using the dependency-dependent and Euclidean distance-based methods introduced in Chapter 4. While the first method remains largely unchanged in this chapter, three new approaches are proposed to accommodate multiple classes in the Euclidean distance-based method. Using two datasets, we extensively tested all possible combinations of relevant features identified using the two methods, and their performances are reported.

5.1 Introduction 38
5.2 Related Work 39
5.3 Problem Statement 40
5.4 Methods 40
5.4.1 Dataset Preparation 40
5.4.2 Feature Selection 41
5.4.2.1 Minimum Distance Approach 41
5.4.2.2 Maximum Distance Approach 41

5.4.2.3 Average Distance Approach 42
5.4.3 Training and Validation 42
5.5 Performance Evaluation 42

5.1 Introduction

In Chapter 4, we introduced the dependency-dependent and Euclidean distance-based feature selection methods. We showed that ML models trained on candidate panels identified using these methods achieve high accuracy in identifying early cases of Alzheimer's disease. However, previous works utilizing blood data for Alzheimer's diagnosis, such as the methods introduced in chapter 4, have primarily focused on identifying biomarker panels to classify Alzheimer's patients from healthy individuals rather than distinguishing between disease stages. In contrast, studies employing MRI data, such as the method discussed in chapter 3, have shown success in classifying different stages of Alzheimer's disease. Using blood data for disease stage classification poses challenges due to the overlap between disease stages. This chapter employs the dependency-dependent and Euclidean distance-based methods for disease stage classification while also addressing some of the limitations of the methodology in chapter 4. For example, this chapter employs a broader panel formation method compared to chapter 4, which could help with the significant drop in performance seen in chapter 4 for set 1 in both feature selection methods.

Previous works in Alzheimer's disease stage classification, besides the method discussed in Chapter 3, include a CNN-based method utilizing MRI scans achieving accuracy rates as high as 97%, an approach combining MRI with clinical data achieving 79.8% accuracy, and a methodology combining MRI, PET, CSF, and genetic information achieving 66.7% accuracy [3, 41, 42]. In contrast, this chapter evaluates the performance of ML models in Alzheimer's disease stage classification when trained on blood data and feature panels formed using the dependency-dependent and Euclidean distance-based methods introduced in Chapter 4. The first method ranks features based on their dependency on the diagnosis, measured using metrics such as MI, SU, and CV, and remains largely unchanged. However, this chapter discusses three new approaches to the Euclidean distance-based feature selection method to accommodate multiple classes. In this method, class means are used to score features. All possible combinations of top-ranked features and features above specified thresholds are generated and used in training models. This chapter is organized as follows: Section 5.2 discusses related works, Section 5.3 presents the problem description, Section 5.4 outlines the methodologies, and Section 5.5 presents the performance of the ML models.

5.2 Related Work

Several studies are currently underway to identify biomarkers that could accurately diagnose Alzheimer's disease [1]. Among them, blood-based biomarkers are popular, as blood tests are regarded as relatively non-invasive and cost-effective compared to CSF tests and imaging solutions such as beta-amyloid imaging using PET or MRI [11]. In Chapter 4, we introduced two novel feature selection methods to identify potential biomarker panels from blood-proteomic data to classify individuals in the early stages of Alzheimer's disease. The proposed methods filtered relevant features based on their dependency on the label, measured using metrics such as MI, SU, and CV, and using the Euclidean distance between class means for each feature. In the chapter, we saw that ML models trained on candidate panels identified using these two feature selection methods achieved as high as 98.56% accuracy.

Several previous works have successfully identified relevant features using blood data and achieved remarkable results in binary classification. However, finding relevant features for multi-class classification of disease stages has been challenging due to the similarities between the three disease stages and requires further development [3]. Many previous works aim to identify features that could help classify patients, regardless of their disease stage, to detect early cases. Typically, in these works, ML models are trained using AD and NL samples and then tested using MCI and NL samples obtained earlier to evaluate the model's performance in early disease detection [11]. Llano et al., on the other hand, utilized univariate analysis to identify features with clear distinctions among AD, MCI, and NL samples [36]. The analysis revealed a clear distinction between the means of the three classes in some features [36]. However, the panels formed using the identified features were used for binary classification, where models were trained and tested similarly to other works [36].

Most previous works on Alzheimer's disease stage classification utilize imaging data, such as the method discussed in Chapter 3. Notably, a CNN-based approach using MRI scans achieved 97% accuracy in classifying samples into four disease stages: non-demented, very mild demented, mild demented, and moderate demented [41]. Another work uses a combination of clinical and MRI data [42]. Clinical data, including Functional Activities Questionnaire (FAQ), Neuropsychiatric Inventory (NPI), and Geriatric Depression Scale (GDS), are combined with texture-based features extracted from MRI to train SVM, KNN, decision tree, and ensemble models [42]. The best performance was achieved using KNN with 79.8% accuracy in classifying samples into three disease stages (AD, MCI and NL) [42]. Additionally, the work demonstrates that SVM models can achieve similar accuracy when trained on CSF regions extracted from the MRI along with the clinical data [42]. The highest accuracy achieved using SVM was 78.6% [42]. Lin et al. propose a similar approach

for multi-class classification that utilizes a combination of MRI, PET, CSF, and the presence of the APOE $\epsilon 4$ gene [3]. Each modality in this approach is scored using LDA and trained on an extreme learning machine-based decision tree model [3]. Three-way classification, consisting of MCI, AD, and NL samples, and four-way classification, consisting of stable MCI (sMCI), progressive MCI (pMCI), AD, and NL samples, was performed and achieved 66.7% and 57.3% accuracies, respectively [3].

5.3 Problem Statement

Blood-based biomarkers offer a relatively non-invasive and cost-effective alternative to CSF tests and imaging solutions such as MRI and beta-amyloid imaging using PET [11]. Previous research employing ML models trained on blood proteomic data has demonstrated high accuracy in identifying individuals in the early stages of Alzheimer's disease. However, existing early detection approaches using blood have predominantly focused on identifying feature panels that differentiate patients with Alzheimer's disease from normal individuals rather than classifying them according to their specific disease stage. While similar classification has been achieved using ML models trained on MRI data, MRI is expensive compared to a blood-based solution and may not be practical for individuals requiring medical devices such as pacemakers [33]. This chapter specifically investigates the performance of ML models in the multi-class classification of Alzheimer's patients into their respective disease stages using two feature selection methods previously successful for early disease detection.

5.4 Methods

5.4.1 Dataset Preparation

This chapter uses 146 protein measurements from 500µl of EDTA plasma samples obtained from over five hundred ADNI patients. Each patient's plasma samples were collected twice: first at baseline and then 12 months from the baseline. The measurements are standardized using z-scores, and two sets of data were formed, similar to Chapter 4. Samples collected 12 months from baseline are used for training, while those obtained at baseline are used for testing in set 1. Conversely, an 80-20 split of the data, irrespective of when the sample is collected, is used for training and testing, respectively, in set 2. Set 1 aims to evaluate the models' ability to capture the differences between NL, AD, and MCI when exclusively trained on data collected from a single period. In contrast, set 2 simulates real-life scenarios where estimating the disease progression may be difficult for each patient.

The dataset comprises 112 NL, 741 MCI, and 209 AD samples. The imbalance in sample

size risks models overfitting to the majority class and, as a result, introduces bias. To address this issue, data augmentation methods such as random oversampling, undersampling, and SMOTE are used. Discretized datasets are then generated from the augmented datasets using the Multi-Interval Discretization method detailed in Algorithm 4.1. Following the augmentation and discretization steps for the two datasets, feature selection is applied across all 12 datasets.

5.4.2 Feature Selection

Measures like Pearson correlation may not be suitable for our dataset since they measure the linear relationship between two continuous values, whereas our labels are discrete. The relevance of the dependency-dependent feature selection method compared to traditional correlation-based approaches is its advantage that the type of variables is irrelevant. In this method, features are ranked based on their degree of association with the labels, determined using MI, SU, and CV. CV is specifically used for the discretized datasets. For each dataset, sets of top n features are iteratively formed by increasing n from 1 to 146. These feature sets are later used in panel formation.

The Euclidean distance-based feature selection method uses the distance between the class means of each feature to identify relevant features based on the principle that similar samples would cluster around their respective class means in lower dimensions. In this method, features are iteratively filtered out using threshold values in 10 equal descending steps, where the minimum and maximum distances among the features are used as the minimum and maximum values for the threshold. Sets of relevant features identified in each iteration are then used for panel formation.

While the Euclidean distance-based method is straightforward to implement in binary classification with only one distance to consider, it becomes more complex in multi-class classification scenarios, where there are n classes and $\frac{n(n-1)}{2}$ distances to consider. In this chapter, we introduce three approaches to address this problem.

5.4.2.1 Minimum Distance Approach

In this strict feature-selection approach, given the set of distances between the class means $\{d_1, d_2, ..., d_n\}$ for a feature, we select the minimum value in the set. This method ensures that even a slight overlap between classes would lead to the feature being readily filtered out by the threshold.

5.4.2.2 Maximum Distance Approach

In this approach, we consider the set of distances between the class means $\{d_1, d_2, ..., d_n\}$ for a feature, and we select the maximum value in the set. Unlike the minimum distance

approach, this method does not penalize features where some classes overlap. However, it has a similar effect as the minimum distance approach for features where all classes overlap.

5.4.2.3 Average Distance Approach

In this approach, we consider the set of distances between the class means $\{d_1, d_2, ..., d_n\}$ for a feature, and we calculate the average of the set. This method introduces a penalty for features where some classes overlap, unlike the maximum distance approach. However, it has a similar effect as the minimum distance approach for features where all classes overlap.

5.4.3 Training and Validation

In Chapter 4, the panels used for validation are the sets of features formed according to the feature selection methods introduced in the chapter. However, despite the models' high accuracy during validation, they failed to meet the minimum performance criteria during testing, especially for set 1. In this chapter, panels are formed using all possible combinations of features from the feature sets formed in the previous section. Duplicate panels formed for each dataset are removed to prevent redundant training and testing. The correlations among the features are calculated, and panels with a correlation of 0.8 or higher are filtered out to reduce redundancy among the features. For the discretized datasets, normalized MI is used instead of correlation. SVM [16] and KNN [15] models are trained on the panels.

Grid Search [20] and 10-fold cross-validation [23] were used to identify the optimal hyperparameters for each panel. The selection criterion used is the weighted f1-score, where the hyperparameter yielding the maximum f1-score was chosen. For SVM, the search used linear and RBF kernels, with C values 0.1, 1, and 10, and γ values 0.01, 0.1, 1, $\frac{1}{\text{panel\_size}*\text{variance}}$, and $\frac{1}{\text{panel\_size}}$. Conversely, for KNN, the search used k values ranging from 1 to 20. In the search process, the F1-score was preferred over accuracy to evaluate model performance, prioritizing each class's true positives to enhance their recall, maximizing patients' accurate classification into their true disease stage. Panels were filtered based on a minimum performance criterion of 75% recall for each class during cross-validation.

Candidate panels that meet the minimum performance criteria undergo testing for 100 iterations. The average recall for each class, accuracy, and reliability are recorded. Fig. 5.1 visually represents the framework discussed in this chapter.

5.5 Performance Evaluation

Over four thousand unique candidate panels were generated using the dependency-dependent method. However, most models tested on these panels failed to accurately classify MCI sam-

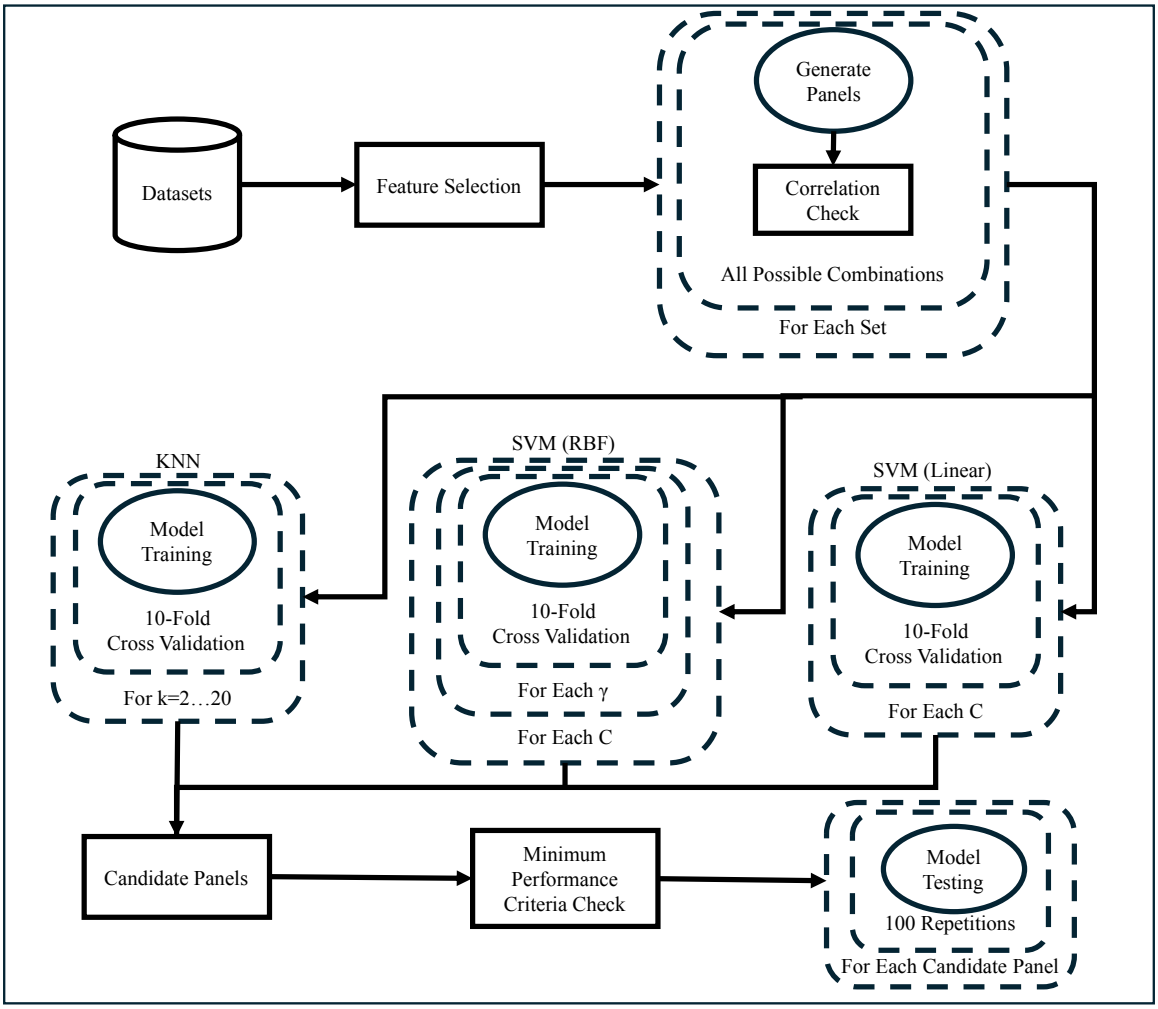

Figure 5.1: Overall Framework for Disease Stage Classification using Blood.

ples, consistently achieving recall as low as 0 for MCI across different panels. None of the panels formed using set 1 met the minimum performance criteria during testing, while 83 panels formed using set 2 met the criteria. Among these, 35 panels were ranked according to MI, 30 according to SU, 17 according to CV, and the last panel contained all the features. These panels were trained on SVM using RBF kernel with C=10 and $\gamma=\frac{1}{\text{panel\_size}*\text{variance}}$, and data was augmented using SMOTE and discretized. A panel of 66 features achieved the best accuracy of 75.17% in set 1, with a recall of 76.25%, 72.22%, and 77.02% for AD, NL, and MCI, respectively. The features in the panel were ranked according to MI and trained on KNN with k=1 using data augmented using SMOTE. Fig. 5.2 illustrates the best recall and accuracy for ten panel sizes during testing leading up to the best performance. The panel containing all the features achieved the highest accuracy of 85.46% in set 2, with a recall of 88.59%, 75.84%, and 91.95% for AD, NL, and MCI, respectively. However, a smaller panel of only 58 features achieved an accuracy of 80.54%, with a recall of 80.54%, 69.13%, and 91.95% for AD, NL, and MCI, respectively. This panel is considered the best performing despite not meeting the minimum performance criteria due to the lower recall for NL. This performance is comparable to the highest-performing panel with less than half of the features. The features in this panel were ranked according to SU and trained on SVM using RBF kernel with C=10 and $\gamma=\frac{1}{\text{panel\_size}*\text{variance}}$. Fig. 5.3 illustrates the best recall and accuracy for six panel sizes leading up to the best-performing and subsequent panel sizes during testing.

Over sixteen thousand unique candidate panels were formed using the Euclidean distance-based method, but most didn't meet the minimum performance criteria, similar to the dependency-dependent method. Interestingly, unlike the dependency-dependent method, the models maintained good recall for MCI but lower recall for NL. None of the panels formed using set 1 met the minimum performance criteria. One panel of 64 features, formed using the maximum distance approach with a threshold value of 0.4, achieved an accuracy of 72.39%, with 72.73%, 64.39%, and 80.05% recall for AD, NL, and MCI, respectively, when trained on KNN with k=1. For set 1, the highest accuracy achieved using the minimum distance approach was 70.79%, and for the average distance approach, it was 70.37%. Fig. 5.4 illustrates the best recall and accuracy for ten panel sizes during testing, leading up to the best-performing panel. For set 2, four panels met the minimum performance criteria. The highest accuracy of 85.23% was achieved by a panel containing 120 features formed using the minimum distance approach with a threshold of 0.01. This panel was trained on SVM with RBF kernel (C=10, γ=0.01) and achieved 86.58%, 80.54%, and 88.59% recall for AD, NL, and MCI, respectively. The best-performing panel for set 2 achieved 80.76% accuracy during testing with 83.22%, 69.13% and 89.93% recall for AD, NL and MCI, respectively, with only 85 features. Despite not meeting the minimum performance criteria due to low NL recall, the panel achieved a slightly lower performance with significantly fewer features

Table 5.1: Comparison of Performances in Alzheimer's Disease Stage Classification Methods [3, 41, 42]

Work	**Approach**	**Accuracy**
Alshammari et al.	CNN + MRI	97.0%
Altaf et al.	MRI Texture + Clinical Data	79.8%
Altaf et al.	MRI CSF + Clinical Data	78.6%
Lin et al.	MRI + PET + CSF + APOE ϵ4 gene	66.7%
Chapter 3	MRI Brain Region Extraction	99.74%
This Chapter	**Dependency-Dependent**	**85.46%**
This Chapter	**Minimum Euclidean Distance**	**85.23%**
This Chapter	**Maximum Euclidean Distance**	**85.46%**
This Chapter	**Average Euclidean Distance**	**85.23%**

than the highest-performing panel. The panel was trained on SVM using RBF kernel with C=10 and $\gamma = \frac{1}{\text{panel\_size} * \text{variance}}$. Fig. 5.5 illustrates the best recall and accuracy during testing.

All the panels formed using the two methods that met the minimum performance criteria achieved 100% reliability during testing. Table 5.1 presents the best performance reported in this chapter, alongside the performances from previous works. The statistical significance of the performances reported in this chapter compared to the best performance among the previous works reported in [41] is evaluated using a t-test. Specifically, the highest accuracies and their corresponding recalls for the dependency-dependent, minimum, maximum, and average distance methods are compared with the previous result. The test showed that the performance is statistically significant, validating the reported performances' reliability. The three methods discussed in this thesis have demonstrated strong and statistically significant performances in disease diagnosis and disease stage classification. The following chapter discusses a potential future direction, laying the groundwork for developing a continuous disease monitoring tool.

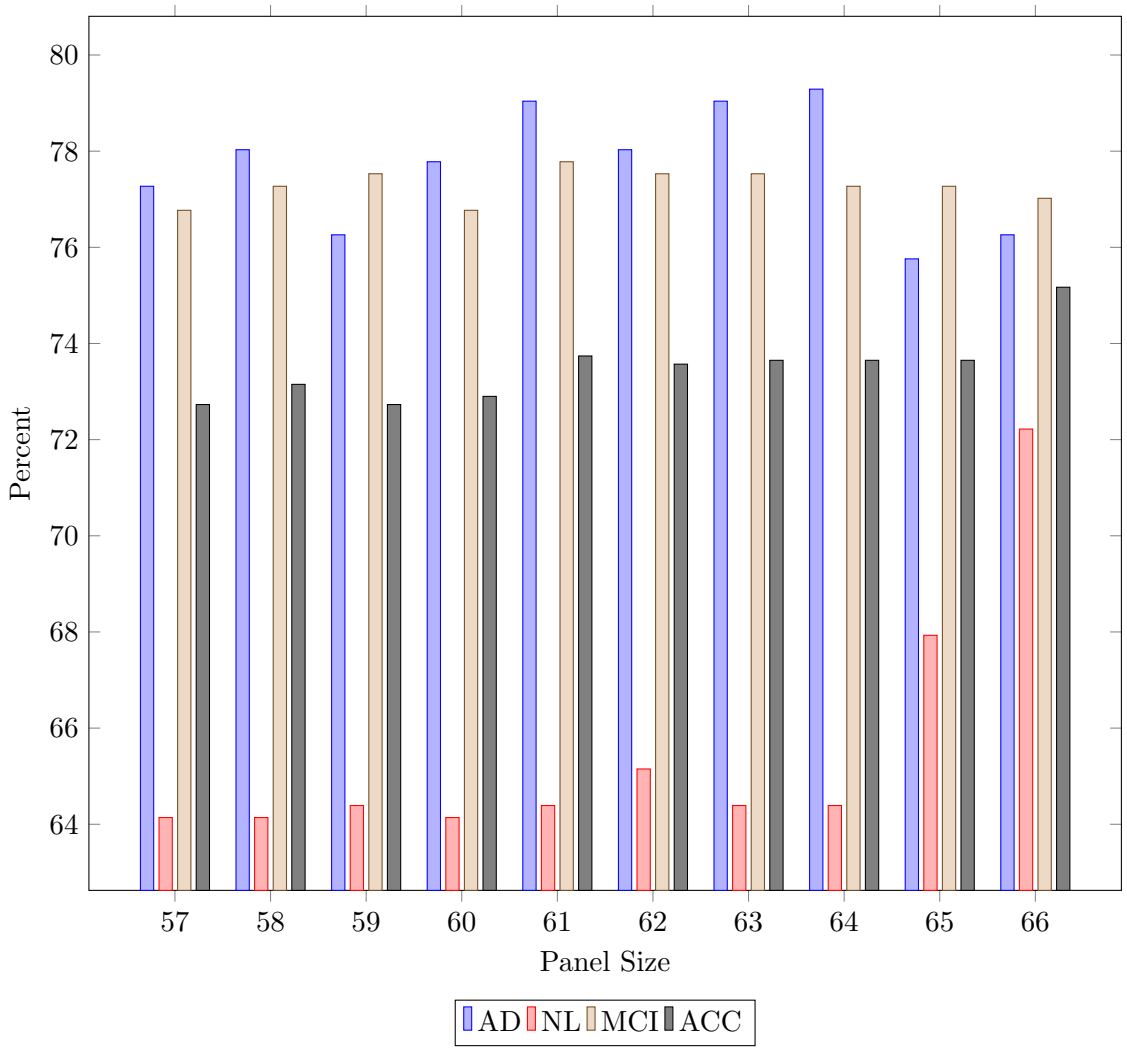

Figure 5.2: Recall and Accuracy for Set 1 (Dependency-Dependent method).

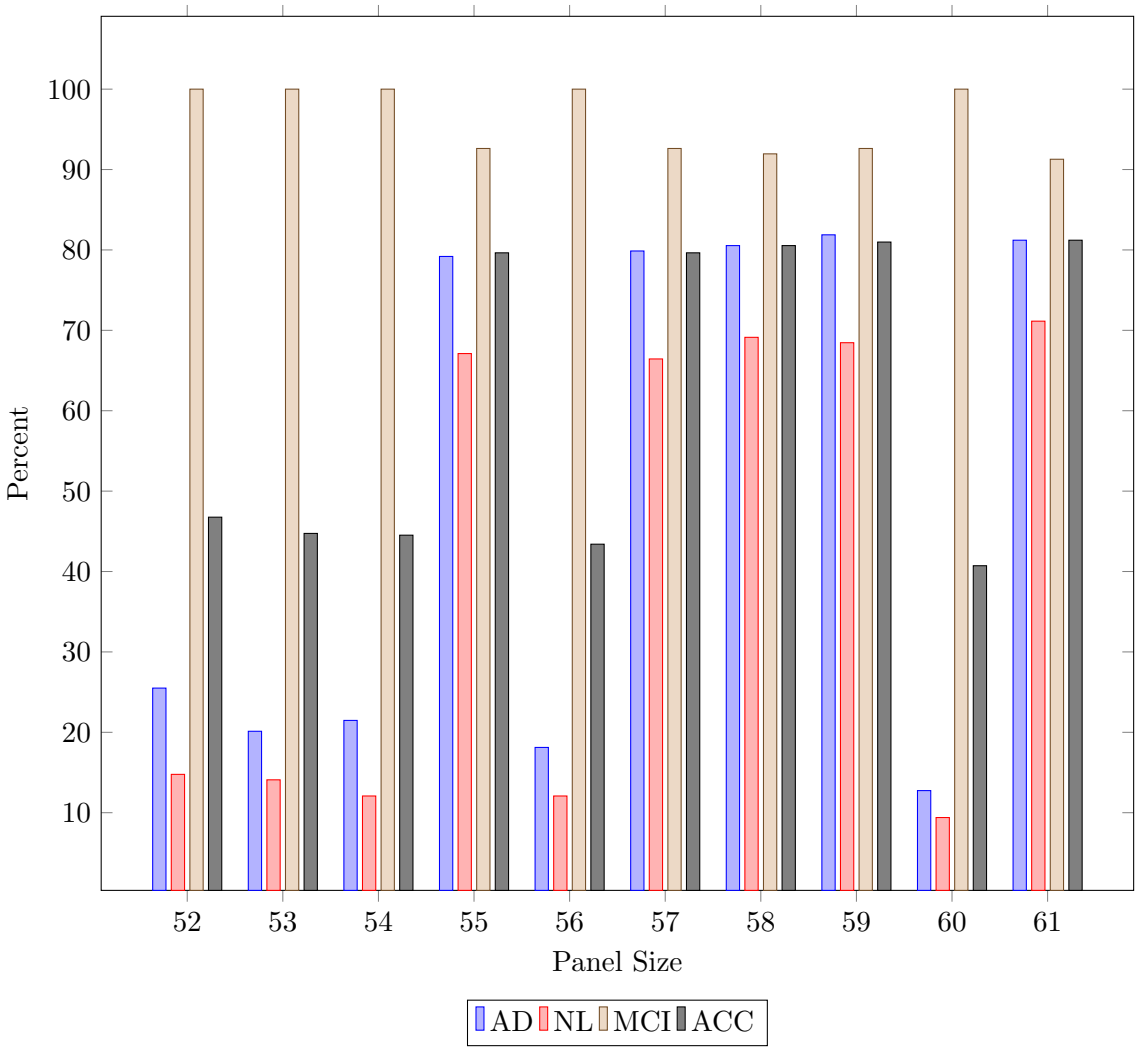

Figure 5.3: Recall and Accuracy for Set 2 (Dependency-Dependent method).

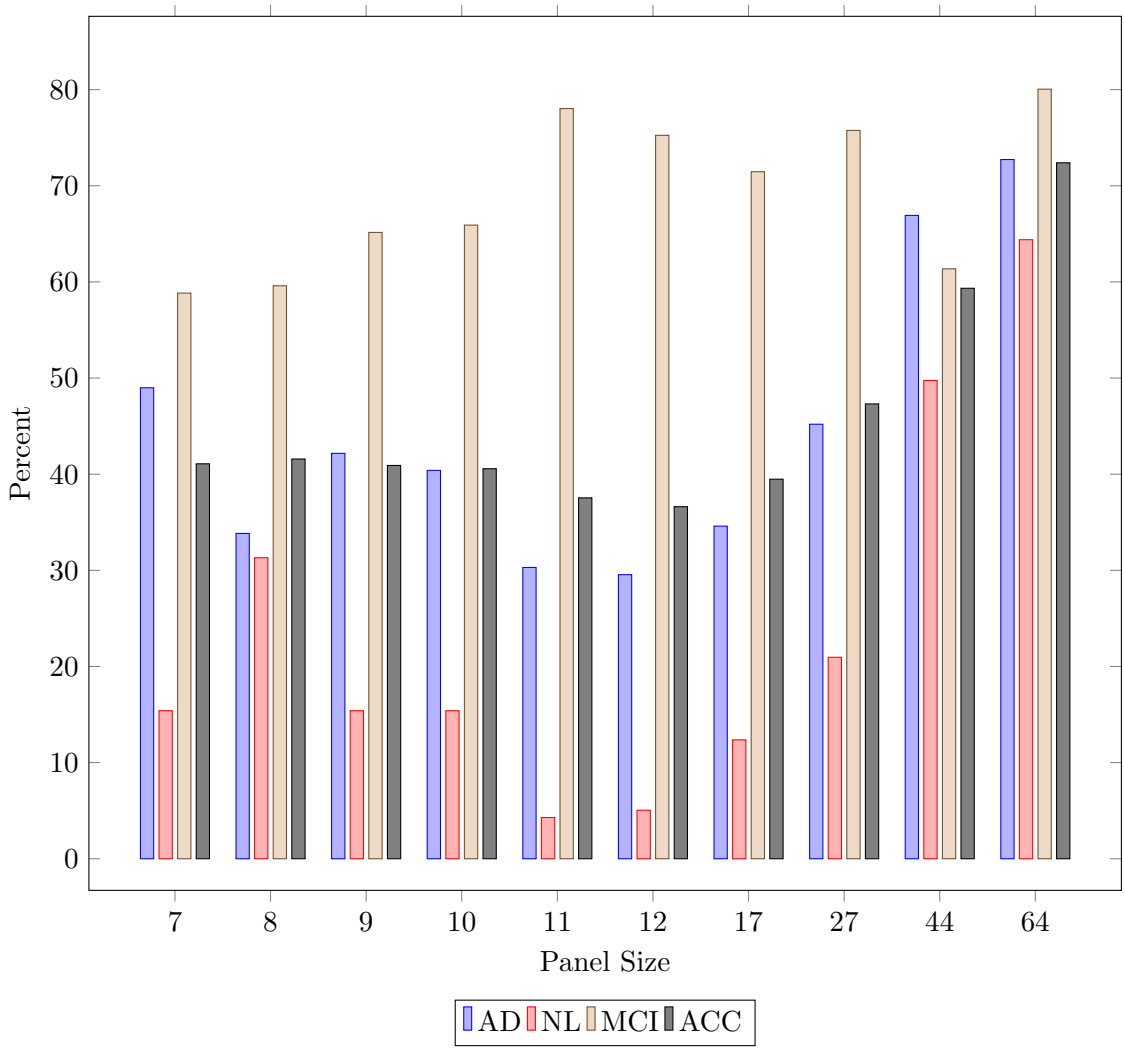

Figure 5.4: Recall and Accuracy for Set 1 (Euclidean Distance-based Method).

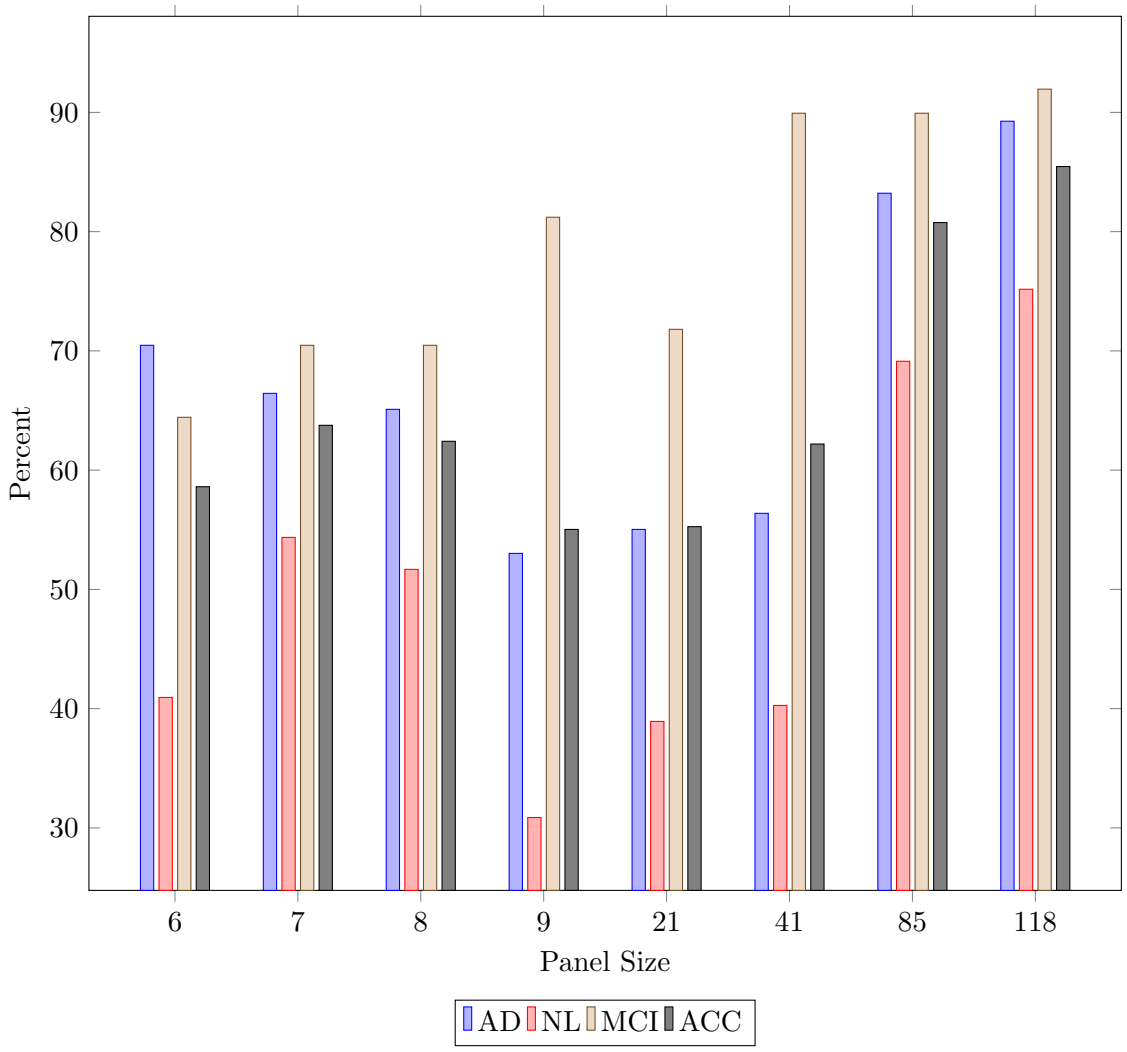

Figure 5.5: Recall and Accuracy for Set 2 (Euclidean Distance-based Method).

Chapter 6

Future Scope: Near Infrared-Based Disease Detection as a Non-Invasive Approach

In Chapter 3, we introduced an imaging-based solution, while Chapters 4 and 5 introduced blood-based solutions for diagnosing Alzheimer's disease, both achieving high accuracies. Despite the advancements in non-invasive and inexpensive solutions that facilitate accessible disease screenings, individuals tend to delay seeking such screenings for various reasons. For example, individuals may attribute Alzheimer's symptoms to normal aging processes, especially among individuals with MCI, where the symptoms may be subtle. This tendency poses a substantial obstacle to early diagnosis. One potential solution is the integration of screening tools into everyday devices. Smartwatches, for example, have shown promise in previous studies aimed at patient screening. A study by Apple and Stanford University, spanning one year and three months, involved collecting and analyzing Electrocardiogram (ECG) data from over 400,000 volunteered Apple Watch users to identify irregular heart rhythms indicative of potentially serious heart conditions [43]. Participants who volunteered for the study received automatic notifications from their smartwatches if irregularities were detected in their ECG readings [43]. Those participants who were notified of irregular heat rhythms underwent further testing, leading to the diagnosis of atrial fibrillation in 34% of cases, while 84% exhibited symptoms consistent with atrial fibrillation [44].

One study used NIR sensors embedded in smartwatches to analyze blood glucose levels, revealing a high correlation between the metabolic index measured by the sensor and blood glucose levels [45]. This study is an important proof of concept of integrating screening tools into everyday devices, especially given that previous research has explored NIR as a potential tool for Alzheimer's diagnosis. A study used NIR spectroscopy to identify changes in cerebral blood and tissue oxygenation during brain activities in AD patients

compared to NL [46]. The study showed distinct patterns of cerebral oxygenation with AD compared to NL [46]. Specifically, AD individuals exhibited a significant decrease in the local concentration of oxygenated hemoglobin and total hemoglobin in the parietal cortex during vocabulary fluency tasks [46]. Furthermore, activation-induced reductions in cerebral hemoglobin oxygenation levels, particularly in the parietal cortex, were observed in AD patients [46].

Another study used NIR spectroscopy to analyze the blood plasma composition of AD, MCI and NL subjects [47]. The results showed values for MCI overlapping with AD and NL subjects [47]. Subsequently, a logistic regression model was trained on the features from the blood plasma analysis for classification between AD and NL subjects and achieved 80% SN and 77% SP [47]. Chou et al. employed NIR to measure activation-induced changes in cerebral hemoglobin concentration, revealing lower mean increases in hemoglobin levels in the frontal cortex among elderly patients during calculation tasks [48]. NIR spectroscopy also detected activation deficits in the parietal lobe of patients with suspected MCI subjects, suggesting its potential as an early diagnostic tool [48]. Furthermore, AD patients exhibited pronounced decreases in hemoglobin levels in the parietal cortex during verbal fluency tasks and a loss of left hemispheric activations during verbal frequency tasks [48]. Finally, Li et al. analyzed changes in oxygenated hemoglobin concentration in different brain regions of AD patients and healthy individuals, highlighting significant differences, particularly in the frontal-parietal areas [49].

Based on the findings of previous studies using NIR for analyzing the characteristics of Alzheimer's patients, we propose the use of NIR-based tools for continuous observation and screening of Alzheimer's, using devices such as smartwatches as a future scope of research.

Chapter 7

Conclusions

The thesis introduces two approaches to Alzheimer's disease diagnosis. The first approach uses volumes of specific brain regions derived from MRI to train traditional ML models. The approach uses a pre-trained CNN-based brain segmentation model for segmenting MRI into five brain regions that have shown abnormalities due to Alzheimer's disease in clinical studies. The volumes of the regions of interest are calculated from the segmented MRI and used to train KNN and SVM models. This approach achieved an accuracy of 99.74% in classifying patients into three classes: AD, MCI and NL. One of the primary objectives of this approach is to eliminate the substantial computational resources typically required for training CNN models for Alzheimer's disease diagnosis. While this objective was partly achieved using a pre-trained model for brain segmentation, which doesn't require significant computational resources, training such a CNN-based model requires significant computational resources. However, in this case, a single model can be used across various applications, eliminating the need to train individual CNN models for each specific application.

The second approach used feature panels identified using blood data for disease diagnosis. Initially, two feature selection methods were used to identify relevant blood features for identifying early cases. The first method ranked features according to their dependency on the diagnosis, measured using MI, SU and CV. The second method identified relevant features using the Euclidean distance between the class means of each feature. Feature panels identified using the two feature selection methods were trained on two sets of data. The first method achieved 90.16% SN and 85% SP, while the second achieved 99.48% SN and 95% SP. These methods were extended to classify patients into three classes: AD, MCI and NL. While the first method remained largely unchained, three new scoring approaches were introduced to accommodate multiple classes in the second method. The scoring approaches involve taking each class's minimum, maximum, and average distances. Panels identified using the two methods were trained on SVM and KNN models. The models achieved accuracies as high as 85.46% using the first method, 85.23% using the minimum distance,

85.46% using the maximum distance, and 85.23% using the average distance approaches.

The proposed approaches are proven to be highly accurate in Alzheimer's disease diagnosis, especially when compared to previous related work. Some of their limitations are discussed in Section 7.1.

7.1 Limitations

Some of the limitations of the methodologies include.

- The approach introduced in Chapter 3 depends on the accuracy of the brain segmentation model. While the pre-trained model used in the chapter is a well-known model that is accurate in segmenting the brain compared to a brain atlas, not all brain segmentation models may have the same accuracy.
- The performance of a generalized brain segmentation model in segmentation may be poor compared to a model specifically designed for the application. For example, the brain segmentation model in Chapter 3 took 3 minutes and 24 seconds on average to segment the brain into 104 regions. A model designed to segment the brain into only the five specific regions of interest would have taken significantly less time.
- The multi-interval discretization algorithm introduced in Chapter 4 may result in some information loss in the dataset as the algorithm may only capture some patterns present in the original continuous features.
- The two feature selection methods introduced in Chapter 4 rely on the classes being balanced. Panels formed on imbalanced datasets might not be accurate.
- The Euclidean distance-based feature selection method introduced in Chapter 4 uses a single threshold value for all the features. However, this might not be a meaningful comparison for two features having different ranges. For example, for two features with ranges (1,2) and (1,4), the threshold could be too large, unfairly filtering out the first feature or too small to accurately filter out the second feature.
- The proposed methodology in this thesis requires a significant amount of time to complete. For instance, segmentation for 1075 samples took over two days for Chapter 3. Additionally, training models in Chapter 4 took over 13 hours, while models in Chapter 5 required almost a full day. The processes were run on an ARM-based processor with 18GB of RAM.

Bibliography

[1] A. A. , "2018 alzheimer's disease facts and figures," *Alzheimer's Dimensia*, vol. 12, no. 3, pp. 367–429, 2018.

[2] V. S. Rallabandi, K. Tulpule, and M. Gattu, "Automatic classification of cognitively normal, mild cognitive impairment and alzheimer's disease using structural mri analysis," *Informatics in Medicine Unlocked*, vol. 18, p. 100305, 2020. [Online]. Available: http://dx.doi.org/10.1016/j.imu.2020.100305

[3] W. Lin, Q. Gao, M. Du, W. Chen, and T. Tong, "Multiclass diagnosis of stages of alzheimer's disease using linear discriminant analysis scoring for multimodal data," *Computers in Biology and Medicine*, vol. 134, p. 104478, Jul. 2021. [Online]. Available: http://dx.doi.org/10.1016/j.compbiomed.2021.104478

[4] A. Juganavar, A. Joshi, and T. Shegekar, "Navigating early alzheimer's diagnosis: A comprehensive review of diagnostic innovations," *Cureus*, vol. 15, no. 9, p. e44937, 2023. [Online]. Available: https://doi.org/10.7759/cureus.44937

[5] M. Chapleau, L. Iaccarino, D. Soleimani-Meigooni, and G. Rabinovici, "The role of amyloid pet in imaging neurodegenerative disorders: A review," *Journal of Nuclear Medicine*, vol. 63, no. Suppl 1, pp. 13S–19S, 2022. [Online]. Available: https://doi.org/10.2967/jnumed.121.263195

[6] S. Chen, J. Zhang, X. Wei, and Q. Zhang, "Alzheimer's disease classification using structural mri based on convolutional neural networks," in *Proceedings of the 2020 2nd International Conference on Big-data Service and Intelligent Computation*, ser. BDSIC 2020. ACM, Dec. 2020. [Online]. Available: http://dx.doi.org/10.1145/3440054.3440056

[7] A. Farooq, S. Anwar, M. Awais, and S. Rehman, "A deep cnn based multi-class classification of alzheimer's disease using mri," in *2017 IEEE International Conference on Imaging Systems and Techniques (IST)*, 2017, pp. 1–6.

[8] K. Aderghal, J. Benois-Pineau, and K. Afdel, "Classification of smri for alzheimer's disease diagnosis with cnn: Single siamese networks with 2d+? approach and fusion on adni," in *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, ser. ICMR '17. ACM, Jun. 2017. [Online]. Available: http://dx.doi.org/10.1145/3078971.3079010

[9] V. S. Rallabandi, K. Tulpule, and M. Gattu, "Automatic classification of cognitively normal, mild cognitive impairment and alzheimer's disease using structural mri analysis," *Informatics in Medicine Unlocked*, vol. 18, p. 100305, 2020. [Online]. Available: http://dx.doi.org/10.1016/j.imu.2020.100305

[10] W. Yang, R. L. Lui, J.-H. Gao, T. F. Chan, S.-T. Yau, R. A. Sperling, and X. Huang, "Independent component analysis-based classification of alzheimer's disease mri data," *Journal of Alzheimer's Disease*, vol. 24, no. 4, p. 775–783, May 2011. [Online]. Available: http://dx.doi.org/10.3233/JAD-2011-101371

[11] C. S. Eke, E. Jammeh, X. Li, C. Carroll, S. Pearson, and E. Ifeachor, "Early detection of alzheimer's disease with blood plasma proteins using support vector machines," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 1, pp. 218–226, 2021.

[12] G. B. Frisoni, N. C. Fox, C. R. J. Jack, P. Scheltens, and P. M. Thompson, "The clinical use of structural mri in alzheimer disease," *Nature Reviews Neurology*, vol. 6, no. 2, pp. 67–77, 2010. [Online]. Available: https://doi.org/10.1038/nrneurol.2009.215

[13] O. L. López and S. T. DeKosky, "Clinical symptoms in alzheimer's disease," in *Dementias*, ser. Handbook of Clinical Neurology. Elsevier, 2008, vol. 89, pp. 207–216. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0072975207012195

[14] F. H. Bouwman, G. B. Frisoni, S. C. Johnson, X. Chen, S. Engelborghs, T. Ikeuchi, C. Paquet, C. Ritchie, S. Bozeat, F. C. Quevenco, and C. Teunissen, "Clinical application of csf biomarkers for alzheimer's disease: From rationale to ratios," *Alzheimer's & dementia (Amsterdam, Netherlands)*, vol. 14, no. 1, p. e12314, 2022.

[15] A. Mucherino, P. J. Papajorgji, and P. M. Pardalos, "k-nearest neighbor classification," in *Springer Optimization and Its Applications, vol. 34*, 2009, pp. 83–106.

[16] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[17] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[18] S. Haykin, *Neural networks: a comprehensive foundation.* Prentice Hall PTR, 1994.

[19] K. Fukushima, "Neural network model for a mechanism of pattern recognition unaffected by shift in position - neocognitron," *Trans. IECE*, vol. J62-A, no. 10, pp. 658–665, 1979.

[20] B. H. Shekar and G. Dagnew, "Grid search-based hyperparameter tuning and classification of microarray cancer data," in *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*, 2019, pp. 1–8.

[21] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[22] A. Fernandez, S. Garcia, F. Herrera, and N. V. Chawla, "Smote for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary," *Journal of Artificial Intelligence Research*, vol. 61, p. 863–905, Apr. 2018. [Online]. Available: http://dx.doi.org/10.1613/jair.1.11192

[23] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-validation," in *Encyclopedia of Database Systems*, 2009, pp. 532–538.

[24] B. C. Ross, "Mutual information between discrete and continuous data sets," *PLoS ONE*, vol. 9, no. 2, p. e87357, Feb. 2014. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0087357

[25] I. Kojadinovic, "On the use of mutual information in data analysis: an overview," in *Proc Int Symp Appl Stochastic Models Data Anal*, 2005, pp. 738–47.

[26] H. Akoglu, "User's guide to correlation coefficients," *Turkish Journal of Emergency Medicine*, vol. 18, no. 3, p. 91–93, Sep. 2018. [Online]. Available: http://dx.doi.org/10.1016/j.tjem.2018.08.001

[27] G. Van Rossum and F. L. Drake Jr, *Python reference manual.* Centrum voor Wiskunde en Informatica Amsterdam, 1995.

[28] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[29] O. Ben Ahmed, J. Benois-Pineau, M. Allard, C. Ben Amar, and G. Catheline, "Classification of alzheimer's disease subjects from mri using hippocampal visual features," *Multimedia Tools and Applications*, vol. 74, no. 4, p. 1249–1266, Jun. 2014. [Online]. Available: http://dx.doi.org/10.1007/s11042-014-2123-y

[30] M. Masoud, F. Hu, and S. Plis, "Brainchop: In-browser mri volumetric segmentation and rendering," *Journal of Open Source Software*, vol. 8, no. 83, p. 5098, 2023. [Online]. Available: https://doi.org/10.21105/joss.05098

[31] A. W. Salehi, P. Baglat, B. B. Sharma, G. Gupta, and A. Upadhya, "A cnn model: Earlier diagnosis and classification of alzheimer disease using mri," in *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, 2020, pp. 156–161.

[32] "ADNI imaging dataset," Data set, 2004. [Online]. Available: https://adni.loni.usc.edu/

[33] M. Ghadimi and A. Sapra, "Magnetic resonance imaging contraindications," *StatPearls*, May 2024, updated 2023 May 1. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK551669/

[34] C. S. Eke, E. Jammeh, X. Li, C. Carroll, S. Pearson, and E. Ifeachor, "Identification of optimum panel of blood-based biomarkers for alzheimer's disease diagnosis using machine learning," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, Jul. 2018. [Online]. Available: http://dx.doi.org/10.1109/EMBC.2018.8513293

[35] E. Jammeh, P. Zhao, C. Carroll, S. Pearson, and E. Ifeachor, "Identification of blood biomarkers for use in point of care diagnosis tool for Alzheimer's disease," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 2415–2418.

[36] D. A. Llano, V. Devanarayan, and A. J. Simon, "Evaluation of plasma proteomic data for alzheimer disease state classification and for the prediction of progression from mild cognitive impairment to Alzheimer disease," *Alzheimer Disease & Associated Disorders*, vol. 27, no. 3, p. 233–243, 2013.

[37] S. E. O'Bryant *et al.*, "A blood-based screening tool for Alzheimer's disease that spans serum and plasma: Findings from TARC and ADNI," *PLoS ONE*, vol. 6, no. 12, article no. e28092, 2011.

[38] H. Soares *et al.*, "Biomarkers Consortium Project: Use of Targeted Multiplex Proteomic Strategies to Identify Plasma-Based Biomarkers in Alzheimer's Disease,"

2010. [Online]. Available: https://adni.loni.usc.edu/wp-content/uploads/2010/11/BC_Plasma_Proteomics_Data_Primer.pdf

[39] U. M. Fayyad and K. B. Irani, "Multi-interval discretization of continuous-valued attributes for classification learning," in *International Joint Conference on Artificial Intelligence*, 1993.

[40] L. Prechelt, "Early stopping — but when?" in *Neural Networks: Tricks of the Trade*, 2012, pp. 53–67.

[41] M. Alshammari and M. Mezher, "A modified convolutional neural networks for mri-based images for detection and stage classification of alzheimer disease," in *2021 National Computing Colleges Conference (NCCC)*, 2021, pp. 1–7.

[42] T. Altaf, S. M. Anwar, N. Gul, M. N. Majeed, and M. Majid, "Multi-class alzheimer's disease classification using image and clinical features," *Biomedical Signal Processing and Control*, vol. 43, p. 64–74, May 2018. [Online]. Available: http://dx.doi.org/10.1016/j.bspc.2018.02.019

[43] M. M. Turakhia and M. V. Perez, "Apple heart study: Assessment of wristwatch-based photoplethysmography to identify cardiac arrhythmias," 2020. [Online]. Available: https://clinicaltrials.gov/study/NCT03335800

[44] M. V. Perez, K. W. Mahaffey, H. Hedlin, J. S. Rumsfeld, A. Garcia, T. Ferris, V. Balasubramanian, A. M. Russo, A. Rajmane, L. Cheung, G. Hung, J. Lee, P. Kowey, N. Talati, D. Nag, S. E. Gummidipundi, A. Beatty, M. T. Hills, S. Desai, C. B. Granger, M. Desai, and M. P. Turakhia, "Large-scale assessment of a smartwatch to identify atrial fibrillation," *New England Journal of Medicine*, vol. 381, no. 20, p. 1909–1917, Nov. 2019. [Online]. Available: http://dx.doi.org/10.1056/NEJMoa1901183

[45] T. Nakazawa, R. Sekine, M. Kitabayashi, Y. Hashimoto, A. Ienaka, K. Morishita, T. Fujii, M. Ito, and F. Matsushita, "Non-invasive blood glucose estimation method based on the phase delay between oxy- and deoxyhemoglobin using visible and near-infrared spectroscopy," *Journal of Biomedical Optics*, vol. 29, no. 03, Mar. 2024. [Online]. Available: http://dx.doi.org/10.1117/1.JBO.29.3.037001

[46] C. HOCK, K. VILLRINGER, F. MÜLLER-SPAHN, M. HOFMANN, S. SCHUH-HOFER, H. HEEKEREN, R. WENZEL, U. DIRNAGL, and A. VILLRINGER, "Near infrared spectroscopy in the diagnosis of alzheimer's diseasea," *Annals of the New York Academy of Sciences*, vol. 777, no. 1, p. 22–29, Jan. 1996. [Online]. Available: http://dx.doi.org/10.1111/j.1749-6632.1996.tb34397.x

[47] D. H. Burns, S. Rosendahl, D. Bandilla, O. C. Maes, H. M. Chertkow, and H. M. Schipper, "Near-infrared spectroscopy of blood plasma for diagnosis of sporadic alzheimer's disease," *Journal of Alzheimer's Disease*, vol. 17, no. 2, p. 391–397, Jun. 2009. [Online]. Available: http://dx.doi.org/10.3233/JAD-2009-1053

[48] P.-H. Chou and T.-H. Lan, "The role of near-infrared spectroscopy in alzheimer's disease," *Journal of Clinical Gerontology and Geriatrics*, vol. 4, no. 2, p. 33–36, Jun. 2013. [Online]. Available: http://dx.doi.org/10.1016/j.jcgg.2013.01.002

[49] R. Li, G. Rui, W. Chen, S. Li, P. E. Schulz, and Y. Zhang, "Early detection of alzheimer's disease using non-invasive near-infrared spectroscopy," *Frontiers in Aging Neuroscience*, vol. 10, Nov. 2018. [Online]. Available: http://dx.doi.org/10.3389/fnagi.2018.00366