

**CLASSIFICATION OF COGNITIVE FRAILTY IN  
ELDERLY PEOPLE FROM BLOOD SAMPLES  
USING MACHINE LEARNING**

**SHAHIDAN BIN IDRIS**

**ELECTRICAL & ELECTRONICS ENGINEERING  
UNIVERSITI TEKNOLOGI PETRONAS**

**MAY 2021**



**Classification of Cognitive Frailty in Elderly People from Blood Samples using  
Machine Learning**

by

Shahidan bin Idris

25204

Dissertation submitted in partial fulfilment of  
the requirements for the  
Bachelor of Engineering with Honours  
(Electrical and Electronics Engineering)

MAY 2021

Universiti Teknologi PETRONAS,  
32610, Bandar Seri Iskandar,  
Perak Darul Ridzuan.

CERTIFICATION OF APPROVAL

**Classification of Cognitive Frailty in Elderly People from Blood Samples using  
Machine Learning**

by

Shahidan bin Idris

25204

Project dissertation submitted to the  
Electrical & Electronics Engineering Programme  
Universiti Teknologi PETRONAS  
in partial fulfilment of the requirement for the  
BACHELOR OF ENGINEERING WITH HONOURS  
(ELECTRICAL & ELECTRONICS ENGINEERING)

Approved by,



Associate Professor Dr Nasreen binti Badruddin

UNIVERSITI TEKNOLOGI PETRONAS

BANDAR SERI ISKANDAR, PERAK

May 2020

### CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

  
Shahidan  
\_\_\_\_\_  
SHAHIDAN BIN IDRIS

## **ABSTRACT**

Cognitive Frailty (CF) is a prevalent age-related disease that is affecting many individuals worldwide. The increasing population worldwide means an increasing elderly population, which is followed by the increase of age-related diseases. Medical intervention needs to be timely, as the late stages of CF prove to be challenging for both clinicians and caretakers. While the existing clinical diagnosis and screening tools for CF are capable of detecting the syndrome, their roles are played mostly in the late-stage detection of CF. As such, a means of prediction is needed in order to identify CF in older adults before its onset. Aside from a means of prediction, key biomarkers for CF are needed to provide pathophysiological clues for medical researchers in developing a standard diagnosis for CF. This study proposed a model to classify patients into different levels of CF, using blood-based parameters from blood samples. The model was also used to identify the top performing biomarkers that are associated with CF. In the study, the dataset from the Malaysian Elders Longitudinal Research (MELoR) was used, which consists of socio-demographic and medical data of older adults. The dataset was cleaned from errors and missing data, followed by pre-processing and input optimisation. Classification tasks were split into 3 parts – multiclass classification of 6 levels of CF, binary classification of Robust and Non-Robust, and binary classification of Robust and Frail with MCI. For the binary classification of Robust and Frail with MCI, 3 different class sizes were used, which were 76 samples, 343 samples, and 100 samples. A total of 7 models using 7 different classification algorithms (Logistic Regression, Linear Discriminant Analysis, k-Nearest Neighbor, Classification & Regression Tree, Gaussian Naive Bayes, Support Vector Machines, and Random Forest Classifier) were trained. The highest accuracy achieved in 6-level classification was 39.0% due to the class imbalance problem and so binary classification was conducted. The highest classification accuracy of 92.0% for binary classification of Robust and Frail with MCI was achieved when 343 samples were used for both classes. From the feature selection, it was identified that the top

ranked features from the classification models have a strong correlation with the potential biomarkers identified by medical researchers and through literature review. While the study has shown promising results in the research on classifying CF using blood samples, future work in this area should focus on ensuring a sufficiently sized dataset with a balanced class distribution is used, as well as sourcing data from a large variety of demographics.

## **ACKNOWLEDGEMENTS**

I would like to take this opportunity to thank the Department of Electrical & Electronics Engineering, Universiti Teknologi PETRONAS for allowing me to undergo the study and providing the necessary logistics support. I would like to also thank my project supervisor, Associate Professor Dr Nasreen Badruddin, for her continuous support, guidance, and motivation throughout the duration of the study. I am grateful to the Ministry of Education, Malaysia, for supporting the study through the Long-Term Research Grant Scheme, Ref: 600-RMC/LRGS 5/3 (002/2019). Special thanks go to Professors Tan Maw Pin, Abu Bakar Abdul Majeed, and Kalavathy Ramasamy for the use of the MELoR dataset and their valuable comments in aiding this study. Last but not least, I would like to thank my family and friends for their encouragement in completing the study.

## TABLE OF CONTENTS

<b>CERTIFICATION OF APPROVAL</b>	<b>ii</b>
<b>CERTIFICATION OF ORIGINALITY</b>	<b>iii</b>
<b>ABSTRACT</b>	<b>iv</b>
<b>ACKNOWLEDGEMENTS</b>	<b>vi</b>
<b>LIST OF FIGURES</b>	<b>x</b>
<b>LIST OF TABLES</b>	<b>xi</b>
<b>CHAPTER 1: INTRODUCTION</b>	<b>1</b>
1.1    Background	1
1.2    Problem Statement	3
1.3    Objectives and Scope of Study	3
<b>CHAPTER 2: LITERATURE REVIEW</b>	<b>5</b>
2.1    Frailty, Mild Cognitive Impairment, and Cognitive Frailty	5
2.1.1    Frailty	5
2.1.2    Mild Cognitive Impairment (MCI)	6
2.1.3    Cognitive Frailty	8
2.2    Diagnosis of Cognitive Frailty	9
2.3    Blood-based Parameters as Biomarkers or Indicators for Cognitive Frailty	10
2.4    Classification of Frailty, MCI, or Cognitive Frailty using Statistical or Machine Learning	13
2.5    Classification Algorithms used in Machine Learning	14
2.5.1    Logistic Regression	15
2.5.2    Linear Discriminant Analysis (LDA)	16
2.5.3    Support Vector Machines	16
2.5.4    Decision Trees	17
2.5.4.1    Classification & Regression Trees (CART)	18

2.5.5	Random Forest	18
2.6	Class Imbalance and Overcoming Class Imbalance	19
2.6.1	Class Imbalance, Causes, and Impact	19
2.6.2	Ways to Overcome Class Imbalance	21
2.6.3	Synthetic Minority Oversampling Technique (SMOTE)	22
2.7	Feature Extraction and Feature Selection	23
<b>CHAPTER 3: METHODOLOGY</b>		25
3.1	Research Methodology	25
3.1.1	Data Collection & Selection	26
3.1.2	Data Pre-processing	27
3.1.3	Input Optimisation	28
3.1.4	Classification	29
3.1.5	Model Validation	31
3.1.6	Performance Evaluation	32
3.1.7	Feature Selection	32
3.2	Tools	33
3.3	Project Milestones and Gantt Chart	33
<b>CHAPTER 4: RESULTS AND DISCUSSION</b>		36
4.1	Multiclass Classification into 6 Classes	36
4.2	Binary Classification of Robust and Non- Robust	39
4.3	Binary Classification of Robust and Frail with MCI	42
4.4	Significant Blood Biomarkers in CF	54
<b>CHAPTER 5: CONCLUSION AND RECOMMENDATION</b>		61
5.1	Conclusion	61
5.2	Recommendation	62

<b>REFERENCES</b>	63
<b>APPENDICES</b>	70

## **LIST OF FIGURES**

Figure 1.1	Population by age group, sex and ethnic group, 2010 – 2019e, Malaysia	2
Figure 2.1	Plot of Logistic Function	15
Figure 2.2	Working of the SVM Algorithm	17
Figure 3.1	Research Methodology Flow Diagram	25
Figure 3.2	Classification Tasks and Class Sample Distributions Used	30
Figure 4.1	Histogram Plots of Feature Value Distribution by Class Membership	38
Figure 4.2	AUC and ROC Curves (Robust and Non-Robust)	41
Figure 4.3	AUC and ROC Curves (Robust and Frail with MCI – 76 samples)	47
Figure 4.4	AUC and ROC Curves (Robust and Frail with MCI – 343 samples)	48
Figure 4.5	AUC and ROC Curves (Robust and Frail with MCI – 100 samples)	49
Figure 4.6	Plot of Accuracy Scores against Majority Class Size	51
Figure 4.7	Plot of AUC against Majority Class Size	52
Figure 4.8	Confusion Matrix and Performance Evaluation for RFC (Robust and Frail with MCI) – 100 samples	54

## LIST OF TABLES

Table 2.1	Operational Measure of Fried Frailty	9
Table 3.1	Distribution of Class Samples	28
Table 3.2	Project Milestones (FYP 1)	34
Table 3.3	Project Milestones (FYP 2)	34
Table 3.4	Gantt Chart (FYP 1)	34
Table 3.5	Gantt Chart (FYP 2)	35
Table 4.1	Holdout Method Performance Evaluation (6 Classes)	36
Table 4.2	5-fold Cross Validation Performance Evaluation (6 Classes)	37
Table 4.3	Holdout Method Performance Evaluation (Robust and Non-Robust)	39
Table 4.4	5-fold Cross Validation Performance Evaluation (Robust and Non-Robust)	40
Table 4.5	Holdout Method Performance Evaluation (Robust and Frail with MCI – 76 Samples)	43
Table 4.6	5-fold Cross Validation Performance Evaluation (Robust and Frail with MCI – 76 Samples)	43
Table 4.7	Holdout Method Performance Evaluation (Robust and Frail with MCI – 343 Samples)	44
Table 4.8	5-fold Cross Validation Performance Evaluation (Robust and Frail with MCI – 343 Samples)	44
Table 4.9	Holdout Method Performance Evaluation (Robust and Frail with MCI – 100 Samples)	45
Table 4.10	5-fold Cross Validation Performance Evaluation (Robust and Frail with MCI – 100 Samples)	46
Table 4.11	Top 10 Ranked Features – 6 Classes	55
Table 4.12	Top 10 Ranked Features – Robust & Non-Robust	56

Table 4.13	Top 10 Ranked Features – Robust and Frail with MCI (76 samples)	57
Table 4.14	Top 10 Ranked Features – Robust and Frail with MCI (343 samples)	58
Table 4.15	Top 10 Ranked Features – Robust and Frail with MCI (100 samples)	59

# **CHAPTER 1**

## **INTRODUCTION**

### **1.1 Background**

The ageing population worldwide is increasing over the years. This trend also holds true for Malaysia, as shown in Figure 1.1. This increase in the elderly population is accompanied by the increase of age-related diseases and disabilities. Two of the common syndromes that affect older adults are frailty and cognitive impairment. In the past, frailty has been studied and defined separately from cognitive impairment. However, recent studies have combined the two syndromes into a single complex phenotype known as Cognitive Frailty (CF). According to an International Consensus Group, CF is defined as “A syndrome in older adults with evidence of both physical frailty and cognitive impairment without a clinical diagnosis of Alzheimer’s Disease (AD) or another dementia” [1].

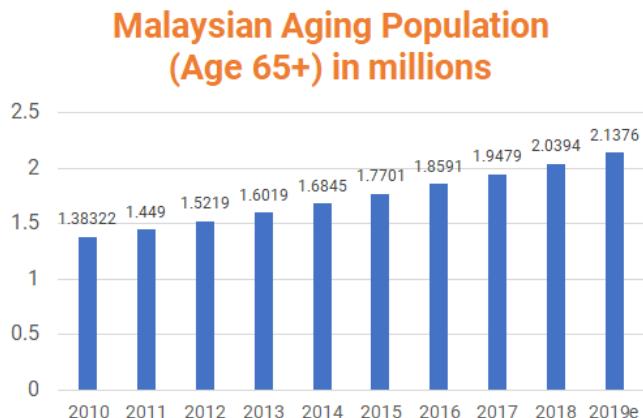


Figure 1.1: Population by age group, sex and ethnic group, 2010 - 2019 (estimated), Malaysia [2]

CF causes much disruption to the lives of not only the patient, but also the people around them. Studies have shown that older adults with CF have significant limitations in carrying out instrumental activities of daily living (IADL). These IADLs include using a bus or train, management of finances, grocery shopping, and housekeeping [3]. There is also research on the association of CF syndrome with mortality, stating that the absolute mortality rate for frail, pre-frail, and normal individuals is 70%, 45%, and 34% respectively [4]. It was found, based on a study conducted, that the prevalence of CF among Malaysians is 37.4% for Cognitive Pre-Frail and 2.2% for Cognitive Frail [5].

Due to the difficulty of reversing the disabilities associated with CF and other age-related diseases, it is important to take preventive action in dealing with age-related diseases. While there are numerous screening tools and measurements for CF and its frailty and cognitive frailty components, these tools take the role of identifying the syndrome during its onset rather than its early or pre-symptomatic stage, which is less than timely. In order to enable proper preventive care and timely medical intervention, research needs to be focused on enabling the prediction of the onset of CF and its symptoms before the syndrome manifests in a patient. These efforts would also aim to identify the various physiological indicators or biomarkers associated with CF, providing medical researchers the ability to draw clinical inferences using the measurements of these biomarkers. In this study, a classification model is proposed to identify CF in

patients using blood samples. The model classifies the disease based on the different blood-based parameters (e.g. red blood cell count, glucose, total cholesterol) into 6 different levels of CF – robust, MCI only, pre-frail only, frail only, pre-frail with MCI, and frail with MCI. From the analysis of the performance of the classification model using different blood-based parameters, the most important features are identified as biomarkers for CF. It is hoped that the results of this study will help medical researchers identify the possible biomarkers associated with CF, in an effort to develop an early, rapid, and standardised clinical diagnosis tool for CF.

## **1.2 Problem Statement**

At the time of writing, most of the existing work aimed at classifying Cognitive Frailty among older adults has been done using questionnaire responses and physical or cognitive assessments results, rather than physiological parameters. Alternatively, the models that predict the condition using physiological parameters have been more focused on physical frailty and mild cognitive impairment (MCI) separately. The attempts made have also mostly conducted clustering (instead of classification) or used statistical analysis techniques rather than modern classification algorithms. Therefore, this project aims to overcome the problem of no existing classification model to predict Cognitive Frailty in elderly people using blood samples.

## **1.3 Objectives and Scope of Study**

The objectives of the project and study are:

- To develop a model to classify different levels of Cognitive Frailty in elderly people using blood-based parameters
- To identify the top features/parameters from blood samples that indicate Cognitive Frailty in elderly people
- To evaluate the effectiveness of the model in terms of accuracy, precision, and F1 score

In terms of the scope of study, the project mainly focuses on:

- The study of CF and its frail and cognitive impairment components, excluding other types of dementia and Alzheimer's Disease
- The blood-based parameters or biomarkers that indicate the different degrees of CF in patients
- The classification of CF among Malaysian older adults living in urban communities
- The classification of CF using the dataset provided from the Malaysian Elders Longitudinal Research (MELoR)

## **CHAPTER 2**

### **LITERATURE REVIEW**

#### **2.1 Frailty, Mild Cognitive Impairment, and Cognitive Frailty**

##### **2.1.1 Frailty**

Frailty is a common condition among older adults which increases the risks of hospitalisation, falls, and morbidity [6]. Many attempts at defining frailty were done by medical professionals and researchers. The various definitions have associated the condition with a patient's disability status [7], functional performance, neurosensory abilities, and comorbidity. While these definitions have associated frailty with different conditions, there was a lack of a definition which measures frailty as a decline across multiple physiological systems, encompassing multiple physiological conditions. An operational definition for the frailty phenotype was presented by Fried and colleagues, which is the definition widely used by medical professionals. Generally, they define frailty in terms of a patient's ability to resist stressors, which degrades as frailty becomes more severe. In their report, Fried defined frailty through 5 phenotypic criteria – weakness, slowness, low level of physical activity, exhaustion and poor endurance, and weight loss [8]. Different degrees of frailty are measured in terms of the number of criteria that a patient fulfils. From the Fried frailty criteria, a patient falls into one of 3 degrees of frailty, which are robust, pre-frail, and frail. There is another definition for frailty, which creates a risk index (known as a Frailty Index or FI) based on the accumulated deficits for various conditions (physical and cognitive impairments, disabilities, diseases, psychological factors, and geriatric symptoms) [9]. Both

definitions have shown to have strong correlation between each other and both definitions have their own strengths and weaknesses. The Fried frailty definition excels in its clinical reproducibility, in which the measurements done to identify the criteria fulfilled can be replicated. Alternatively, the FI definition is better in terms of its mathematical properties, in which clinical inferences can be drawn from such properties [10]. The FI definition is a more sensitive predictor for the condition, as it has a finer graded risk scale. This definition is also more robust compared to the Fried definition, due to the large number of conditions covered by FI. Ultimately, it is up to the medical practitioner or researcher to decide on which definition is more suitable for their work, and that they should clearly specify the definition that is applied when describing CF.

Given the strengths and weaknesses of each definition, this project utilises the Fried definition in its understanding and measurement of frailty. One of the reasons for applying Fried definition over FI is due to the increasing consensus that frailty involves many different symptoms, as such a phenotypic view is preferred. Another reason for using this definition is that the Fried definition is simpler to reproduce in a clinical setting, describing the frailty phenotype in 5 components, as opposed to FI which encompasses 30 to 70 different components. This makes the Fried definition much more appealing in a clinical setting, which greatly benefits the project in terms of its future work and results.

### **2.1.2 Mild Cognitive Impairment (MCI)**

MCI is a condition characterised as cognitive decline that is greater than normal for an individual's age and education level but does not significantly interfere with everyday activities. Throughout the years, there have been a number of attempts made at clinically defining MCI. One of the older attempts was made by Prichard, who defined MCI as impairment of recent memories, while keeping intact distant memories [11]. Kral provided a definition that contradicts Prichard, describing MCI as being forgetful to unimportant information, in which most of this forgotten data comes from the distant

past [12]. A detailed and recent operational definition was provided by Peterson and colleagues. They define MCI as a condition in which a patient self-reports a memory problem and abnormal memory functioning, while having normal activities of daily living (ADLs), general cognitive function, and failure to meet criteria for dementia [13]. In short, a patient with MCI is similar to control subjects, aside from having a memory performance similar to individuals with early Alzheimer's Disease (AD). This definition has been further extended to include the nonamnestic subtype of MCI as well as other cognitive domain subtypes. Furthermore, the self-report (or informant-report) of MCI for a patient needs to be accompanied with an objective cognitive performance measurement, while indicating the patient still has normal functioning of ADLs.

There is no gold standard for a cognitive assessment that diagnoses MCI in a patient. While there are a number of cognitive screening tools available for clinicians, these tools provide measurements in varying and limited numbers of cognitive domains. Two of the more common cognitive screening tools used are the Montreal Cognitive assessment test (MoCa) and Mini Mental state Examination (MMsE). Out of the two, MMsE is more commonly used compared to MoCa. MMsE is a 30-point questionnaire based assessment aimed at measuring cognitive impairment. The tool measures a patient in terms of their orientation, attention, memory, language skills, and visual-spatial skills [14]. Majority of the scores for the tool should be computer-generated to ease statistical analysis. One of the strengths of MMsE is that the administration of the assessment does not require prior training or specialised equipment. Using MMsE also takes up a short amount of time and is convenient to be administered, meaning it can be conducted at a clinic office or bedside. Aside from MMsE, MoCa is also widely used for assessing mild cognitive impairment. Similar to MMsE, MoCa is also a 30-point examination consisting of 11 categories assessing multiple cognitive domains, such as memory, abstraction, language, attention, and visual-spatial functioning [15]. The administration of MoCa takes a long time compared to MMsE, with MoCa taking up 10 to 12 minutes to complete, while MMsE takes up 7 to 8 minutes. One of the strengths of MoCa in terms of providing initial screening for MCI is that MoCa is a more sensitive tool compared to MMsE. Numerous studies have been conducted in comparing both tools to find that

MoCa is better at identifying minor cognitive deficits, whereas MMsE used for patients with MCI tends to give inaccurate results (some patients with MCI were able to score perfectly in MMsE) [16]. As is the same for many screening tools for many diseases and conditions, the usage of both tools strongly depends on what the clinician seeks to screen. If a patient shows only a minor cognitive deficit in early assessments, the clinician should administer MoCa. Alternatively, if a patient shows severe cognitive impairment or symptoms of dementia, there is no need for such a sensitive screening tool like MoCa.

Having said that, the use of cognitive screening tools still should not qualify as standard diagnosis tools. These tools merely provide a severity rating scale for the cognitive deficits of a patient, which is not synonymous to the cognitive impairment syndrome [17]. When discussing MCI, however, it is more viable to administer MoCa as the appropriate screening tool, due to its sensitivity. As such, the project implements the MoCa score as the screening tool to identify patients with MCI.

### **2.1.3 Cognitive Frailty**

In the past, the frailty and cognitive impairment phenotypes have been studied separately. However, numerous studies have found a strong correlation between these two conditions. Patients diagnosed with cognitive impairment or Alzheimer's Disease (AD) have been found to exhibit symptoms of physical frailty [18], [19], while physical frailty symptoms have been found to manifest before the onset of dementia [20], [21]. In 2013, the International Academy on Nutrition and ageing (I.A.N.A.) and the International Association of Gerontology and Geriatrics (I.A.G.G.) organised an International Consensus Group on cognitive frailty (CF), which discussed the condition as a single complex phenotype. The panel also provided the first definition of CF, which is, "a syndrome in older adults with evidence of both physical frailty and cognitive impairment without a clinical diagnosis of AD or another dementia". This definition

extends the existing frailty definition by Fried, incorporating the assessment of cognitive function.

## 2.2 Diagnosis of Cognitive Frailty

Since CF is a combination of frailty and cognitive impairment into a single phenotype, it is therefore plausible to clinically diagnose CF in terms of its frailty and cognitive impairment components. As discussed previously, frailty is defined either in terms of the Fried frailty definition or the Frailty Index (FI) definition. For the project, the Fried definition, thus the Fried measurement of the frailty phenotype, is employed. The Fried measurements of frailty are done according to weakness, slowness, low level of physical activity, exhaustion and poor endurance, and weight loss. Table 2.1 describes the operational measure of frailty for each of the components. The detailed explanation on the criteria used for the definition is provided in Appendix A.

Table 2.1: Operational Measure of Fried Frailty

Fried Frailty Criteria	Measurements
Weakness	Grip Strength, lowest 20% (based on sex, BMI)
Slowness	Walking time/15 ft, slowest 20% (by sex and height)
Low-level of Physical Activity	kcal/week, lowest 20% (by sex)
Exhaustion and Poor Endurance	Self-reported exhaustion
Weight Loss	Unintentional, $\geq 10$ lbs in prior year

According to the above measurements and criteria, the possible levels of frailty for a patient are robust, pre-frail, and frail. If a patient fulfills none of the criteria, the patient is considered robust. If a patient fulfills 1 or 2 criteria, they are considered

pre-frail. If a patient fulfills 3 criteria and more, they are considered frail. As discussed before, the cognitive screening tools are used to identify MCI in a patient. The selected screening tool, due to its higher sensitivity towards mild cognitive deficits, is MoCa. A sample of the MoCa test in English is provided in Appendix B. Since MoCa is a score-based screening tool (perfect score is at 30), a cutoff point for the MoCa score determines whether a patient is normal or has MCI. According to MoCa, if a patient scores 23 to 30, they are considered normal. If a patient scores 0 to 22, they are considered as having MCI.

Combining the diagnoses for frailty and MCI, we are able to identify different levels of CF for a patient. Given the 3 possible conditions for frailty and 2 possible conditions for MCI, we thus have a total of 6 different levels of CF, which are robust, MCI only, pre-frail only, pre-frail with MCI, frail only, and frail with MCI. The project aims to classify CF in patients according to these levels.

### **2.3 Blood-based Parameters as Biomarkers or Indicators for Cognitive Frailty**

Using the clinical diagnosis methods for CF allows a clinician to identify the phenotype within a patient. However, while the definitions and methods employed are able to detect certain conditions relating to CF, there is still no method of providing researchers an idea of how the disease progresses through different levels (from robust to frail with MCI). In a clinical research standpoint, the lack of indicators or biomarkers associated with CF provides a challenge for measuring the severity and progression of the disease in a patient. In recent years, much research has been done in identifying numerous biomarkers for frailty and cognitive impairment, providing a pathophysiological measure for CF.

One of the more popular biomarkers for frailty are inflammatory markers. Studies have shown that the ageing process closely relates with an acceleration in inflammatory activity in the body [22]. 3 of the inflammatory markers that are most

researched are the Interleukin 6 (IL-6), C-Reactive Protein (CRP), and Tumor Necrosis Factor Alpha (TNF $\alpha$ ). IL-6 and TNF $\alpha$  are cytokines, which are proteins secreted by cells related to the body's immune system. These proteins help in regulating inflammation. As the body's inflammatory activity increases, the presence of these cytokines increases, which is supported by numerous studies that show a correlation between the presence of these cytokines and frailty [23], [24]. CRP is a biomarker present in body serum or the bloodstream. The level of CRP increases when there is inflammation in the body. Similar to IL-6 and TNF $\alpha$ , CRP is strongly related to the frailty syndrome [25], [26]. While using inflammatory markers as indicators does potentially translate to the identification of frailty, these markers are still unable to distinguish between different levels of the disease. As such, it is not viable to utilise inflammatory biomarkers alone to measure frailty. Aside from inflammatory markers, the presence of frailty has been shown to cause metabolic irregularities in a patient's body. A common disease associated with this problem is diabetes mellitus (DM). Similarly, a number of researchers have shown that the physiological and neurological impact of DM are precursors to the onset of frailty in a patient [27 - 29]. When studying DM, it is natural to analyse the glycated haemoglobin (HbA1c) in the bloodstream. The HbA1c test is a common measurement of whether DM is present in a patient. The test measures the levels of glucose in the blood plasma that is encountered by the red blood cells in a 3-month period. If the overall average plasma glucose level is high, the glucose levels binded to the haemoglobin increases, providing an accurate measure of DM. Similar to inflammatory markers, while there is a strong correlation between DM and frailty, thus HbA1c and frailty, the measurement of HbA1c is unable to distinguish between different levels of frailty. There is potential for using blood or serum-based biomarkers to measure for frailty. A hypothesis was presented on the systematic irregularity in protein production and metabolism activity due to frailty, which is detectable in the blood. While there are a large number of serum biomarkers that have been studied, some of the key biomarkers for frailty are albumin level, haemoglobin level, and glomerular filtration rate. Studies on the albumin level, circular haemoglobin level, and glomerular filtration rate have indicated their strong correlation with frailty [30 - 33]. In terms of a symptomatic standpoint, these serum biomarkers are

strongly associated with muscle weakness and exhaustion, which are key symptoms of frailty.

There are numerous developments in the screening and measurement of cognitive impairment. Most attempts of screening for MCI, AD, and different types of dementia are done using imaging techniques, such as magnetic resonance imaging (MRI), single-photon emission computed tomography (SPECT) screening, and positron emission tomography (PET) [34]. While these technologies provide ways of studying the changes in structure and fluids of the brain, the establishment of biomarkers is important for the prediction of cognitive impairment. Biomarkers for identifying cognitive impairment are mostly either based on cerebrospinal fluid (CSF) or blood. A few potential biomarkers for cognitive impairment in CSF include  $\beta$ -amyloid peptides (1-42), Tau levels, and amyloid-Tau ratio [35], [36]. The measurement of CSF biomarkers does show high correlation with cognitive impairment, however these biomarkers are less economical and show low interpretability. The assays for collecting CSF biomarkers require specialised equipment and technology, which adds up to the lab and diagnosis costs. The screenings for these biomarkers also require specialised memory clinics. Alternatively, blood-based biomarkers would be more economical and easier to screen. One of the blood-based biomarkers for cognitive impairment is clusterin. Clusterin in the blood plasma has been shown to be altered when there is neurodegeneration [37]. Higher levels of clusterin have been found in patients with MCI compared with patients with normal cognitive function. High levels of clusterin have also been associated with minor brain atrophy in MCI patients. There are a number of studies on plasma lipids as a potential biomarker for MCI. A study by Yin and colleagues identified an inverse relationship between triglyceride levels and MCI [38]. This means that triglycerides have protective qualities against cognitive impairment in a patient. Another study analysed the total cholesterol, high-density lipoprotein cholesterol (HDL-cholesterol), low-density lipoprotein cholesterol (LDL-cholesterol), and triglycerides in patients with MCI [39]. It was found that patients with MCI had lower levels of HDL-cholesterol and plasma triglycerides, which could prove their function in preventing cognitive impairment. Misfolded or unfolded proteins, namely the p53 protein, have also been

found in increased levels in patients with MCI or early AD compared to controls. This could prove unfolded p53 as another potential biomarker for the syndrome [40].

## **2.4 Classification of Frailty, MCI, or Cognitive Frailty using Statistical or Machine Learning**

Using the various screening tools and biomarkers for CF and its frailty and cognitive impairment components, there are numerous attempts at using statistical learning and machine learning approaches to identify high-performing biomarkers and predict the syndrome among patients. One of the attempts involved a clustering analysis on older adults who were robust, pre-frail, and frail. This study utilised the General Practice (GP) electronic health records (eHRs) and patient self-reports, which are a total of 261 records having 10 numerical variables. The study implemented the Elbow Method in identifying the number of clusters for the analysis. As for the clustering, the K-Means Clustering algorithm was used. From the study, a total of 3 clusters were formed, with Cluster 2 showing the highest non-frail patients, while Cluster 3 showing the highest pre-frail patients [41]. While this attempt does not clearly identify or predict the disease in patients through the variables used, this approach does provide clues on potential biomarkers that were deemed useful in the clustering, which are fasting glucose, cholesterol, LDL-cholesterol, glomerular filtration rate, haemoglobin and haematocrit. Some of these biomarkers are in agreement with the previous section about biomarkers for CF. Another study attempted classification and prediction of AD using plasma-signalling proteins. The study made use of self-collected data, consisting of 259 plasma samples and 120 signalling proteins as variables. Rather than using modern machine learning algorithms, this study employed statistical learning approaches in its classification. The clustering analysis was done using significance analysis of microarrays (SAM), while the classification algorithm was predictive analysis of microarrays (PAM). Validation was done using both holdout method (50-50 train-test split) and 10-fold cross validation. From the classification, PAM achieved a 90% positive agreement and 88% negative agreement. Furthermore, 18 types of proteins in

the blood were found to be capable of identifying AD [42]. One study which closely resembles the project for this report has built a predictive model for frailty. The study used data from the Toledo Study for Healthy Aging (TSHA). The dataset had 474 participants with 284 health-related parameters encompassing questionnaire answers, blood parameters, and vitals parameters. The study first did data preparation using Variable Deletion with a missing value ratio of 33.33%, while conducting Multiple Imputation by Chained Equations (MICE) imputation. Feature selection was done using the Boruta Algorithm, and the input data was normalised. The classification algorithms used were Naïve Bayes, Classification And Regression Trees (CART), bagging CART, C5.0, Random Forest, Support Vector Machine (SVM), and Linear Discriminant Analysis (LDA). The classification was validated using 10-fold cross validation. From the results, the SVM model achieved  $78.31 \pm 0.70\%$  accuracy, while the Random Forest model achieved  $77.46 \pm 0.45\%$  accuracy [43]. While this study has classified the frailty syndrome with a satisfactory accuracy score, it was found that the top features that helped in the model performance came from questionnaire answers that were closely related to the frailty and MCI clinical diagnosis tools, rather than physiological parameters. As such, there is still much to do in terms of using machine learning to predict CF using biological parameters, such as parameters from the blood.

## 2.5 Classification Algorithms used in Machine Learning

The machine learning algorithms used for this project consist of supervised learning algorithms. Supervised learning algorithms are machine learning algorithms used on data that have input variables ( $X$ ) along with its respective labels ( $y$ ). In other words, the work of the machine learning algorithms is to map a function onto the dataset, building a relationship between the input variables with the labels. For this project, the machine learning algorithms proposed are Logistic Regression, Linear Discriminant Analysis, K-Nearest Neighbour, Classification and Regression Trees, Gaussian Naive Bayes, Support Vector Machines, and Random Forest.

### 2.5.1 Logistic Regression

Logistic Regression is an algorithm that uses a logistic function to map the relationship between the input variables and the label. This type of algorithm is mostly used for binary classification, where prediction from the input variables would yield one of two outcomes. The basis of the Logistic Regression algorithm lies in the use of the logistic function or sigmoid function shown in Equation 1.

$$y = \frac{1}{1+e^{-x}} \quad (1)$$

The probability of a binary outcome is calculated using the function and an S-shaped logistic graph is produced as shown in Figure 2.1.

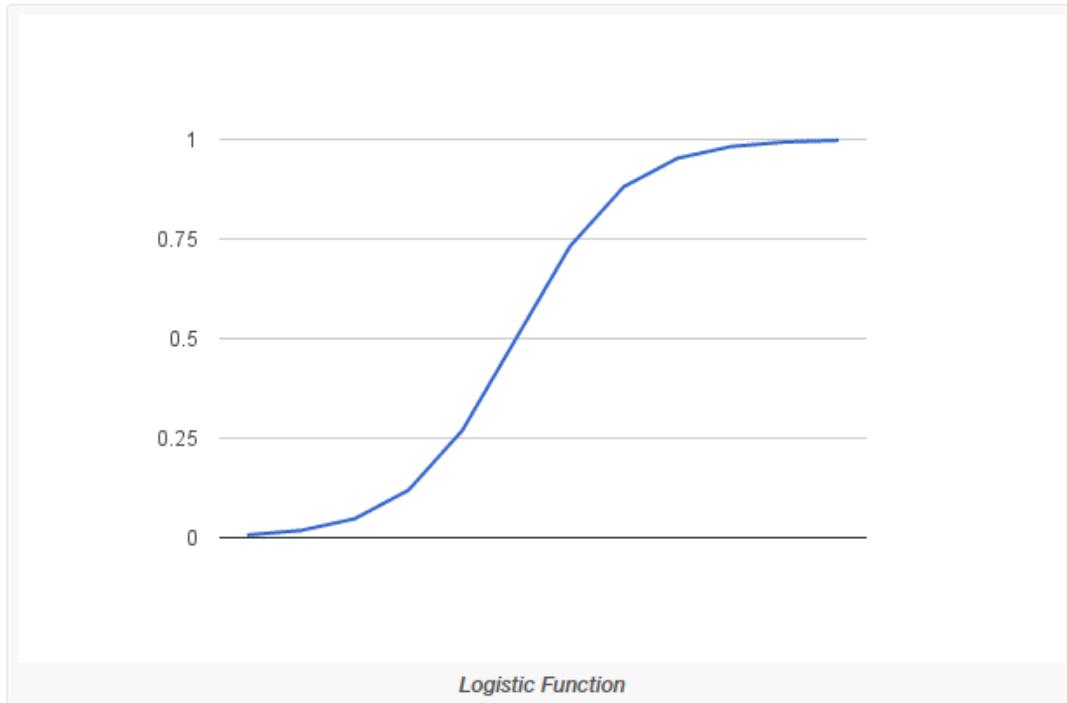


Figure 2.1: Plot of Logistic Function

While Logistic Regression provides an estimate of a sample belonging to one of two classes, the algorithm is best used when the label variables are dichotomous. In the case of multi-class classification, this algorithm may not perform the best.

### 2.5.2 Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis or LDA is another supervised learning classification algorithm in machine learning. The algorithm works by combining the variables of a dataset in a way that the differences between the classes are maximised. LDA assumes equal covariance among the classes. The probability that a new set of inputs belongs to each class is calculated. The output class with the highest probability is chosen. To estimate probabilities, LDA employs Bayes' Theorem. The Bayes Theorem Formula is provided in Equation 2.

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)} \quad (2)$$

$A, B$	=	events
$P(A B)$	=	probability of A given B is true
$P(B A)$	=	probability of B given A is true
$P(A), P(B)$	=	the independent probabilities of A and B

### 2.5.3 Support Vector Machines

Both linear and non-linear data can be categorised using the Support Vector Machine (SVM) algorithm. Each data sample is first mapped into an n-dimensional feature space, where n denotes the number of features. A hyperplane dividing the samples into two classes is defined, with the marginal distance for both classes maximised and classification errors minimised. The marginal distance between the

hyperplane and the nearest instance of that class is the marginal distance for that class. Figure 2.2 shows the working of the SVM algorithm.

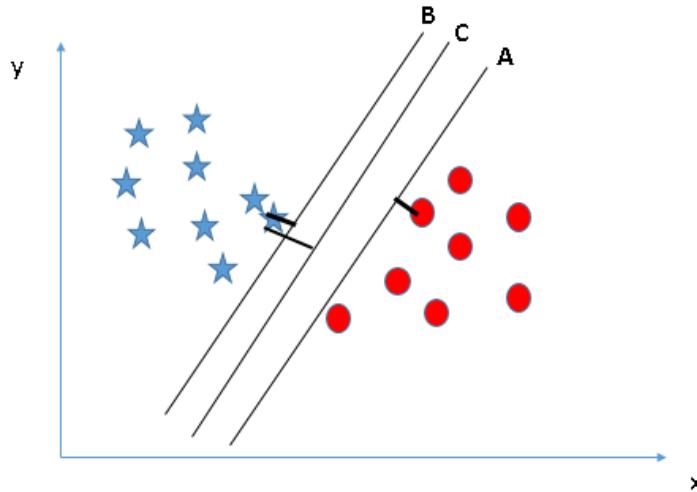


Figure 2.2: Working of the SVM Algorithm

#### 2.5.4 Decision Trees

A decision tree represents the decision logic for classifying data samples into a tree-like structure. The tree's nodes usually have several levels, with the root node being the first or top-most node. All internal nodes (those with at least one child) represent input variable or attribute tests. The classification algorithm branches towards the appropriate child node based on the test result and the process of testing and branching repeats until it reaches the leaf node. The decision outcomes are represented by the leaf or terminal nodes. Decision Trees have been found to be simple to interpret and quick to learn. When traversing the tree for a sample's classification, the results of all tests at each node along the path will provide enough information to make a guess about the sample's class.

#### 2.5.4.1 Classification & Regression Trees (CART)

CART was introduced by Breiman in 1984 [44]. It is an improvement over CHAID. CART constructs binary trees, where each internal node has two outgoing edges. When comparing different Decision Tree algorithms, the binding equations used for each algorithm differ. CHAID uses chi-square tests to find the most dominant feature. Later algorithms after CART, such as ID3 and C4.5, use information gain and gain ratio respectively. As for CART, the algorithm uses a metric known as the GINI index. The equation for the GINI index is provided in Equation 3.

$$Gini = 1 - \sum_{j=1}^C p_j^2 \quad (3)$$

where  $p_j$  = Probability of an item being in class  $j$

$C$  = Total number of classes

A key strength of CART over CHAID is its ability to generate regression trees. This means that the trees have leaves that predict real numbers rather than a class. In a regression task, CART finds splits that minimise prediction squared error. The prediction is done based on the weighted mean for the node.

#### 2.5.5 Random Forest

The Random Forest (RF) classifier is an ensemble classifier made up of many Decision Trees (like a forest made up of many trees). Overfitting of the training data is common with deep Decision Trees, resulting in a large variance in classification outcome for a small change in the input data. They are extremely sensitive to their training results, making them vulnerable to making mistakes with the test dataset. Different sections of the training dataset are used to train the different Decision Trees of

a Random Forest. To identify a new sample, the sample's input vector must be passed down with each Decision Tree of the forest. The classification outcome is then determined by each Decision Tree considering a particular part of the input vector. The forest then decides whether the classification with the most 'votes' or the average of all trees in the forest should be used. Since the algorithm takes into account the effects of several different Decision Trees, it can reduce the variance generated by only considering one Decision Tree for the same dataset.

## **2.6 Class Imbalance and Overcoming Class Imbalance**

### **2.6.1 Class Imbalance, Causes, and Impact**

Class imbalance is a condition whereby the number of samples for each class of a dataset is not evenly distributed. This means, in the case of a binary classification (two classes in the dataset), one of the classes has significantly more samples than the other class. When describing the nature of the class imbalance, a ratio of the minority class size to the majority class size is used. For example, a dataset with an imbalance ratio of 1:100 indicates that for every one sample of the minority class, there are 100 samples of the majority class. When it comes to determining whether a dataset is considered imbalanced, or whether the dataset is slightly or severely imbalanced, there is no fixed imbalance ratio to reflect this scenario. However, most work done in handling the class imbalance problem focuses on imbalance ratios ranging from 1:4 to 1:100 [45]. This serves as a good benchmark to determine if a dataset in question is imbalanced, while being of great use for the study in determining the degree of imbalance of the dataset used. Since class imbalance is determined to be between the range of 1:4 and 1:100, there is a question of what ratio of the majority class to minority class serves as the most well-balanced dataset, when looking at ratios that are lower than 1:4. There is once again no fixed ratio to determine whether a dataset is completely balanced, since the nature of the dataset and values does not guarantee that a 1:1 class distribution produces an absolutely balanced dataset. However, with the lack of ruling or sufficient literature

review that covers the ‘perfect balance ratio’, a logical choice for a class distribution to achieve good class balance is by creating a 1:1 ratio.

Class imbalance is a common problem faced when conducting classification and regression tasks, especially when using real-life datasets. This problem happens even more frequently when handling medical-related data. This is because medical data predominantly consists of ‘normal’ samples, with only a small percentage of the data consisting of ‘abnormal’ samples [46]. In most cases, the classes with the least number of samples (also known as the minority class) tend to be the more important class that is critical to the machine learning task, like predicting a disease [47].

When carrying out prediction tasks, the use of conventional machine learning algorithms requires extreme care with regards to the class distribution. When classifying a sample from a dataset into one of two classes, the machine learning algorithm needs to determine the probability or score of a class membership. When this probability exceeds a certain threshold, known as the ‘decision threshold’, the algorithm classifies this sample within this class. A large number of machine learning algorithms have a default value of 0.5 for their decision thresholds when it comes to a prediction or scoring range from 0 to 1. Using this default threshold value, the machine learning algorithm predicts each sample with the assumption that the class distribution is balanced. In the event that a severe class imbalance is present, the skew or bias in the dataset causes the prediction probabilities to vary. Using the default threshold in this scenario would not be an optimal interpretation.

When using an imbalanced dataset for classification, it is harder for the algorithm to predict samples into the minority class. With the small number of samples, it is more difficult for the classification model to learn the characteristics of the samples and differentiate these samples from the majority class. Furthermore, the large difference in class sizes may cause the majority class to ‘swamp’ the minority class. Referring to the previous discussion on decision thresholds, since machine learning algorithms assume an

even class distribution, the models may naively focus on learning the characteristics of the majority class alone, while ignoring the minority class.

### 2.6.2 Ways to Overcome Class Imbalance

There are two approaches that can be taken to handle the class imbalance problem. The first approach is to control the decision threshold of the classification algorithms. This technique is known as threshold-moving. Threshold-moving helps to accommodate a broader range of the classification problem, while providing an optimum interpretation of the prediction possibilities, as opposed to the default 0.5 threshold value. Using the receiver operating characteristic (ROC) curve, the predicted probabilities of the models are analysed. The best threshold is then identified by searching a range of threshold values, either by determining the geometric mean (G-mean) of the predictions or by using Youden's J statistic [48].

Another simpler approach is by directly manipulating the distribution of the class samples. This is done either by oversampling the minority class or undersampling the majority class. Oversampling is the process of adding artificially generated or duplicated samples to the minority class, while undersampling involves eliminating samples from the majority class. An approach for conducting undersampling is by selecting a subset of the majority class samples randomly (random undersampling). Aside from random undersampling, there are different approaches that have been investigated by researchers to conduct undersampling while preserving the information from the original dataset. One of these techniques is cluster-based undersampling [49]. By using cluster-based undersampling, the dataset is first clustered into  $k$  number of subsets, then the appropriate training samples are taken from the subsets or clusters.

Compared to undersampling, oversampling increases the number of samples in the dataset and prevents information from being lost in the original dataset, since no samples were removed. Oversampling can be done either by duplicating samples or by

using an oversampling approach known as Synthetic Minority Oversampling Technique (SMOTE) [50]. When using oversampling with replacement, a number of researchers have determined that this approach does not significantly improve the recognition of the minority class [51], [52]. The approach of duplicating samples in the minority class only increases the number of samples without considering the impact on the decision region of the feature space. In other words, the replication of samples causes the decision regions to be very specific and not spread into the majority class region, leading to overfitting. An alternative to this approach is by synthesising new samples from the minority class instead of replicating the samples by using SMOTE.

### **2.6.3 Synthetic Minority Oversampling Technique (SMOTE)**

SMOTE works by first identifying the  $k$  number of neighbours surrounding a randomly selected sample in the class. Out of the  $k$  number of neighbours, one of the neighbours is randomly selected. A new sample is then synthesised by creating a new data point between the two randomly selected points in the feature space. Compared to replicating samples, SMOTE creates new samples that are plausible, whereby they are closely related to the existing samples of the class and are close with the existing samples in the feature space. A downside to using SMOTE is that the oversampling does not take into consideration the majority class samples, causing the oversampling and resulting classification to be ambiguous when there is an overlap between the classes.

While SMOTE does increase the model accuracy of the minority class, it is not a good idea to overuse SMOTE by generating too many synthetic samples. This can result in the entire dataset becoming too ‘artificial’, making it not a realistic representation of the original dataset or domain. While the generated models may perform well when evaluated, these models would not be able to perform well when used for real applications. As recommended by the inventor of SMOTE, the best approach to using SMOTE is by combining both undersampling and SMOTE in a dataset. The technique involves undersampling the majority class to a certain percentage of the class size, then

proceeding to conduct SMOTE on the minority class. This technique has been proven to create better performing models that are more suitable for real applications compared to using SMOTE alone [50].

While combining undersampling and SMOTE does produce the best results, there is no fixed number of samples or percentage of the dataset in which SMOTE should be applied. While different combinations of class distributions and undersampling-oversampling ratios were attempted by the author, the performance of each test case with regards to the combinations used were greatly dependent on the nature of the dataset. Since there is no fixed rule on the number of samples to be SMOTE-ed, there is a need to analyse iteratively the impact of the number of SMOTE-ed samples to the performance evaluation scores of the models trained from the dataset used. A good assumption is also to conduct SMOTE for a small percentage of the dataset [50]. This percentage would need to be further verified in terms of the performance evaluation scores and impact towards the confusion matrices generated.

## 2.7 Feature Extraction and Feature Selection

In most medical data, there is an extremely large number of features for each patient or sample. These features take the form of the many symptoms identified in the patient, various physiological parameters (blood glucose level, cholesterol, white blood cell volume etc), as well as demographic or assessment test related answers. The high dimensionality of medical data introduces problems to machine learning tasks, such as classification and regression. The large number of features and the subsequent large size in the dataset introduces increasing problems to both machine learning practitioners and healthcare professionals when it comes to data storage. In terms of the machine learning workflow, the high dimensionality further complicates data analysis and preprocessing, while introducing additional computational cost (time and computing resources) to carry out the classification or regression. As such, a large amount of research has been focused

on dimensionality reduction techniques. Two general approaches to reducing the data dimensionality are feature selection and feature extraction.

Feature extraction is conducted to reduce data dimensionality. The process of feature extraction conducts a transformation of the features into a lower dimensional feature space. This process preserves most of the information of the features, which is an advantage over feature selection. However, since feature extraction involves a linear representation of the original features , it is not possible to interpret the extracted features as compared to selected features. In other words, the information of how much the original features contribute to the prediction is mostly lost. A well-known feature extraction technique is the principle component analysis (PCA).

Aside from feature extraction, it is also possible to conduct feature selection. Feature selection is done by taking a subset of the features present in the dataset. The selection of features are based on the relative importance of these features towards the models, whereby the best subset is the one with both the least number of dimensions and the most contribution towards the learning accuracy of the models [53]. The advantage of feature selection is that information that is contained in a single feature is maintained throughout the process, while providing better interpretability of the selected features. One of the feature selection techniques used is Recursive Feature Elimination (RFE).

One of the algorithms used in feature selection is Recursive Feature Elimination (RFE). The process of RFE involves wrapping the algorithm with a machine learning algorithm to select the subset of features. The algorithm starts with fitting all the features in the dataset. The performance evaluation is done on the fitting to determine how well the model performs using the entire dataset. Each feature is then ranked in terms of importance, with the least important feature discarded. The process repeats by refitting the remaining features, ranking the importance of each feature and eliminating the least important feature. The iteration stops until a certain number of features remain (as determined by the machine learning practitioner) [54].

## CHAPTER 3

### METHODOLOGY

#### 3.1 Research Methodology

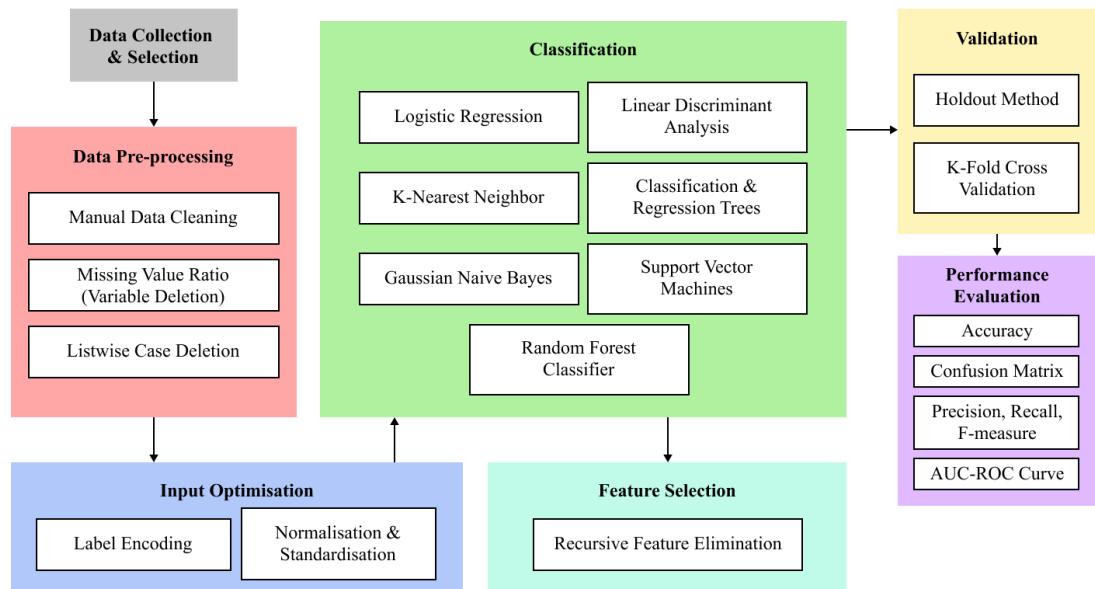


Figure 3.1: Research Methodology Flow Diagram

Figure 3.1 shows the research methodology for the project. The work for the project is separated into 7 parts, which are the data collection & selection, data pre-processing, input optimisation, classification, validation, performance evaluation, and feature selection. Each step is explained in Sections 3.1.1 – 3.1.7:

### **3.1.1 Data Collection & Selection**

The dataset for the project is obtained from the Malaysian Elders Longitudinal Research (MELoR) study. MELoR is a national study conducted by the University of Malaya (UM) to investigate the present challenges faced by the elderly community in Malaysia, with the aim of effectively preparing for the escalating number of the elderly population [55]. This multidimensional study, covering various aspects of ageing, consists of socio-demographic and medical data from participants aged 55 years and above collected since 2013. The participants for the study are urban community-dwelling older adults who were randomly identified from the Malaysian electoral roll at the Petaling Jaya South, Petaling Jaya North and Lembah Pantai parliamentary constituencies. The sampling of participants was stratified by age deciles (5-year age groups) and ethnicity (Malay, Chinese, and Indian) [56]. The data collection was done by trained researchers who attended participants' homes. Questionnaires were first administered to the participants, asking about socio-demographic information and general health. The participants then attended a health check conducted for further health assessments (including physical, physiological, cognitive measurements) and biochemical screenings. The individuals provided written consent before participating in this study [57], [58].

This dissertation focuses on the blood-based parameters, Montreal Cognitive Assessment (MoCA) score, and Fried Frailty data collected in the MELoR study. There are a total of 1652 samples in the MELoR dataset, with 1123 participants having the complete required data needed for the dissertation. From the MoCA scores and Fried Frailty data of each sample, the participants are classified into 6 different groups of CF (Robust, Prefrail only, Frail only, MCI only, Prefrail + MCI, and Frail + MCI), based on the evaluation criteria covered in the Literature Review (Diagnosis of Cognitive Frailty). A new variable, named ‘condition’, is added to denote the 6 groups. From the original 712 variables, 59 variables are retained, which are the patient ID (named ‘mtag’), the ‘condition’ variable, as well as 57 blood-based parameters. The list of retained variables are provided in Appendix C. The dataset after selection has 1123 instances and 59

variables, saved in a comma-separated values (CSV or .csv) file named ‘rawfile\_blood.csv’.

### 3.1.2 Data Pre-processing

As is common for many real-life datasets, there are samples that could have missing or incorrect values. This is due to the process of data acquisition being prone to errors [59]. Given the dataset used for the dissertation is relatively small, it is possible to manually clean the data. In order to handle value errors, the dataset values are cross-checked with the values listed in the data dictionary Word file. Values that are scaled uniformly (e.g. every value in a single variable is scaled by 109) are scaled back to the allowed ranges as per the data dictionary. This process of data cleaning and its final values are then validated by domain experts, consisting of medical professionals and researchers. While this method of data cleaning works well for uniformly scaled variable values, there are columns of data that may not have consistent scaling. In other cases, there may be a mismatch of data types between values of a single variable (e.g. variables with values of ‘>60’ instead of an exact numerical value). Due to the difficulty and inability to retrace and correct these values, any variable facing this issue will be entirely removed from the dataset.

After handling errors, the dataset is then processed for missing values (or NaNs). There are numerous techniques in handling missing values, which are generally separated into two categories – deletion and imputation [60]. While imputation is oftentimes preferred over deletion (because more data will be available for model training), imputation only works best when there are less than 10-25% missing values in a single variable [61]. The dataset is first checked for variables that contain too many NaNs. A missing value ratio of 20% is used, meaning if a variable were to have more than 20% NaNs, the variable would be removed from the dataset. The final list of variables after the numerous cleaning techniques are provided in Appendix D. After variable deletion, the Listwise Case Deletion approach is used. If any case (participant)

has NaNs in any of its variable values, the entire case is removed from the dataset. The finalised dataset after cleaning, with 1015 samples and 48 variables, is then saved into a CSV file named ‘rawfile\_blood\_parsed.csv’.

### 3.1.3 Input Optimisation

Before optimisation, the dataset is first split according to its label (y-variable or dependent variable) and features (x-variables or independent variables). The ‘mtag’ variable is first dropped from the dataset as it does not serve any purpose in the model training. The ‘condition’ variable is taken as the label, while the remaining variables are taken as the features. The distribution of class samples is shown in Table 2.

Table 3.1: Distribution of Class Samples

Class	Class Size
Robust	343
Prefrail with MCI	233
Prefrail Only	223
MCI Only	133
Frail with MCI	76
Frail Only	7

Optimisation of the dataset values is important as machine learning models rely on good data to perform well [62]. Input optimisation processes dataset values into an optimal format before being used for model training. Two techniques are used to carry out optimisation for the feature values, namely standardisation and normalisation. Standardisation is the process of transforming the dataset to have a mean,  $\bar{x}$ , of 0 and standard deviation,  $\sigma$ , of 1. The process approximates the variable to a normal variable.

Using standardisation helps in creating compatibility and similarity between values of a variable, while keeping errors to a minimum. This is particularly useful when the values of different variables vary largely (e.g. one variable with range 0 to 1 and another variable with range 0 to 1000), creating a bias during the model training. Normalisation is the process of casting the values of a variable between a specified range (e.g. between 0 and 1). The differences in ranges of values are not distorted in this process. Datasets that do not contain outliers benefit from this approach [63 - 6]. For label input optimisation, the variable values are in string (text) format. Label encoding technique is used to convert the string values to numerical values. Using this technique, distinct label values are encoded to unique integers (0 for Robust, 1 for Prefrail Only, 2 for Frail Only etc). This converts the data to a machine-readable form, allowing the model to better understand the label values.

### **3.1.4 Classification**

The classification tasks are divided into 3 groups – Multiclass Classification of 6 degrees of CF, Binary Classification of Robust and Non-Robust, and Binary Classification of Robust and Frail with MCI. For the binary classification of Robust and Frail with MCI, 3 different distributions of class samples are used. The classification tasks and class sample distributions used are illustrated in Figure 3.2.

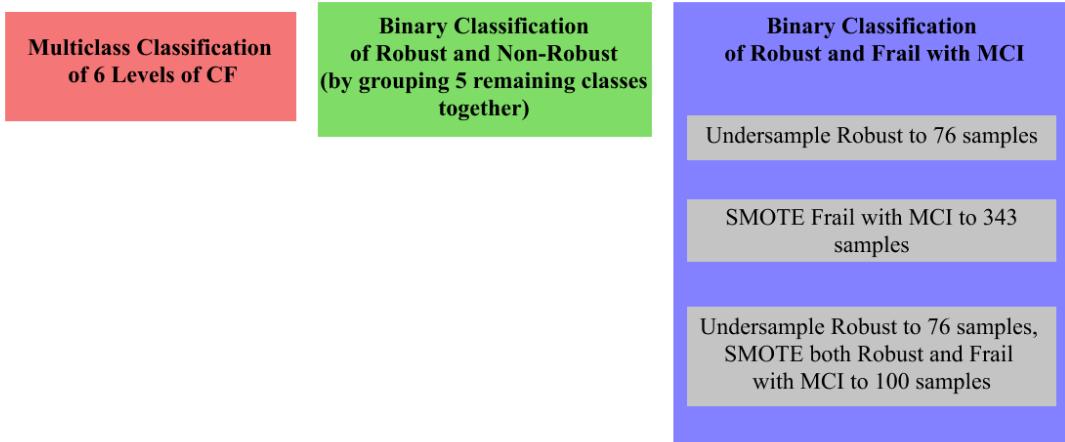


Figure 3.2: Classification Tasks and Class Sample Distributions Used

For the classification of 6 degrees of CF, the entire dataset with the respective classes are used. For the binary classification for Robust and Non-Robust, the 5 remaining classes (aside from Robust) are grouped into a single class ‘Non-Robust’. This creates a new class distribution of Robust with 343 samples and Non-Robust with 672 samples. In order to ensure a well-balanced dataset is used for the classification, the majority class (Non-Robust) is randomly sampled for 343 instances. This creates a 1:1 distribution between the two classes.

For the binary classification of Robust and Frail with MCI, this classification task takes the extreme ends of the CF syndrome spectrum (classifying the most healthy and the most unhealthy) as the dataset. The distribution of class samples for Robust is 343 samples, with Frail with MCI having 76 samples. Within this classification task, different distributions of class samples are used. For the first distribution, the same approach as the binary classification of Robust and Non-Robust is used, whereby the majority class (Robust in this case) is randomly sampled for 76 instances. This creates a 1:1 distribution between the two classes. The second distribution involves oversampling the dataset using the Synthetic Minority Oversampling Technique (SMOTE) to achieve the similar 1:1 distribution. The minority class (Frail with MCI) is oversampled from 76

samples to 343 samples to match the number of samples of Robust. The third distribution is a combination of both the first and second distribution. The majority class (Robust) is first randomly sampled for 76 instances to match Frail with MCI. After achieving the 1:1 distribution, both classes are oversampled using SMOTE to achieve 100 samples for each class. The use of the third distribution allows for sufficient data samples from each class to be used for model training.

7 different machine learning algorithms are used to classify CF using blood-based parameters. The algorithms used are Logistic Regression, Linear Discriminant Analysis, k-Nearest Neighbor, Classification & Regression Tree, Gaussian Naive Bayes, Support Vector Machines, and Random Forest Classifier. A total of 7 models are trained and produced, each trained with one of the algorithms listed. Validation and performance evaluation is done on the 7 models.

### **3.1.5 Model Validation**

Two validation methods are used for the project – holdout method and k-fold cross validation. The 7 models are first trained and validated using the holdout method, which splits the dataset into training and test sets. The models trained using the training set are validated using the test set to measure their performance. The ratio used for the train-test split is 60-40 (60% for training, 40% for test). For cross validation, the full dataset is split randomly into k subsets, with (k - 1) sets used for training and the remaining set used for testing. The process is repeated with the subsets taking turns to be the test set (other subsets used for training). The number of iterations, k, used for the project is 5 (5-fold cross validation).

### **3.1.6 Performance Evaluation**

Each model produced from the 7 algorithms is first evaluated in terms of their accuracy score. In order to fully evaluate the performance of the models, additional measures are used for evaluation. The confusion matrix plot is produced for each model, showing the number of actual vs predicted samples for each class. From the confusion matrix, the precision, recall, and f-measures are calculated. The same evaluation procedures are done for k-fold cross validation, whereby a confusion matrix and subsequent calculations are done for each iteration of k for each model. At the end of the model training, the receiver operating characteristic (ROC) curve is plotted for each model, with the corresponding area-under-curve (AUC) calculated.

### **3.1.7 Feature Selection**

To determine the best features or biomarkers that best help in the classification of CF in patients, the Recursive Feature Elimination (RFE) algorithm is used. From the 7 algorithms used to train classification algorithms, the Logistic Regression, Linear Discriminant Analysis, Classification & Regression Tree, Support Vector Machines, and Random Forest Classifier models are used as wrappers for Recursive Feature Elimination (RFE). From 46 features used in the original dataset, the top 10 features ranked in importance are selected as a result from the RFE. Each model is evaluated in terms of model accuracy to determine the best subset of features as biomarkers. A box and whisker plot is produced, showing the distribution of accuracy scores for each model wrapped with a classification algorithm.

### **3.2 Tools**

- 1. Python (Anaconda Distribution)**

Programming language used to conduct classification of CF using different machine learning algorithms.

- 2. Jupyter Notebook**

Computational notebook to write code, produce computational output, and add code descriptions. Each version of the written software is done in individual notebooks.

- 3. Scikit-learn**

Python library that contains machine learning and predictive analysis tools. The various machine learning algorithms and model validation and evaluation tools were provided by scikit-learn.

- 4. Imbalanced-learn**

Python library that specifically deals with class imbalance, providing different tools and algorithms that aim to solve this problem. SMOTE was used from imbalanced-learn.

- 5. Matplotlib**

Python library that allows for numerical visualisation through graph plotting. Rather than producing separate plot windows from the software, the integration of Jupyter Notebook and matplotlib allows for graphs to be embedded in notebook entries.

### **3.3 Project Milestones and Gantt Chart**

Table 3.2 and Table 3.3 show the project milestones for FYP 1 and FYP 2 respectively. Table 3.4 and Table 3.5 illustrate the Gantt Chart for FYP 1 and FYP 2 respectively.

Table 3.2: Project Milestones (FYP 1)

Milestones/Week	1	2	3	4	5	6	7	8	9	10	11	12
Finalise dataset collection and confirmation of dataset to be used												
Confirm features to be extracted for classification												
Confirm algorithms to do classification												
Complete data pre-processing												
Complete dataset building (train/validate/test)												

Table 3.3: Project Milestones (FYP 2)

Milestones/Week	1	2	3	4	5	6	7	8	9	10	11	12
Complete model training												
Complete model evaluation												

Table 3.4: Gantt Chart (FYP 1)

Tasks/Week	1	2	3	4	5	6	7	8	9	10	11	12
Conduct research and literature review												
Finalise dataset collection and confirmation of dataset to be used												

Identify suitable features for model training from dataset													
Identify appropriate algorithm to do classification													
Data cleaning, verification, and formatting (pre-processing)													
Building and splitting of dataset (training/validation/test)													
Model training and refinement													

Table 3.5: Gantt Chart (FYP 2)

Tasks/Week	1	2	3	4	5	6	7	8	9	10	11	12
Model training and refinement												
Model evaluation												

## CHAPTER 4

### RESULTS AND DISCUSSION

#### 4.1 Multiclass Classification into 6 Classes

The models produced from the different classification tasks (6 classes, Robust and Non-Robust, Robust and Frail with MCI) and class sample distributions (76 samples, 343 samples, 100 samples) were evaluated. Table 4.1 and Table 4.2 show the performance evaluation for the multiclass classification for both holdout method and k-fold cross validation.

Table 4.1: Holdout Method Performance Evaluation (6 Classes)

Algorithms	Accuracy	Precision	Recall	F-Measure
Logistic Regression (LOG)	38.2%	17%	22%	19%
Linear Discriminant Analysis (LDA)	35.0%	18%	21%	19%
K-Nearest Neighbor (KNN)	35.0%	17%	20%	17%
Classification and Regression Tree (CART)	27.3%	20%	21%	20%
Gaussian Naive Bayes (GNB)	36.7%	17%	22%	19%
Support Vector Machines (SVM)	38.2%	16%	21%	17%
Random Forest Classifier (RFC)	37.7%	16%	21%	18%

Table 4.2: 5-fold Cross Validation Performance Evaluation (6 Classes)

Algorithms	Accuracy	Precision	Recall	F-Measure
Logistic Regression (LOG)	38.8% (std. dev. 0.05)	24.2%	22.2%	19.0%
Linear Discriminant Analysis (LDA)	36.3% (std. dev. 0.03)	20.6%	21.4%	19.6%
K-Nearest Neighbor (KNN)	31.4% (std. dev. 0.03)	18.6%	18.2%	16.4%
Classification and Regression Tree (CART)	26.5% (std. dev. 0.02)	18.4%	19.2%	19.0%
Gaussian Naive Bayes (GNB)	33.1% (std. dev. 0.03)	22.6%	22.4%	20.8%
Support Vector Machines (SVM)	39.0% (std. dev. 0.04)	15.8%	21.6%	16.6%
Random Forest Classifier (RFC)	35.6% (std. dev. 0.04)	25.6%	21.2%	17.8%

For the classification of CF into 6 classes, LOG and SVM had the highest holdout method accuracy of 38.2%. For cross validation, SVM showed the highest accuracy of 39.0%. While the accuracy scores for the models were higher than the probability of classifying a sample into one out of the 6 classes ( $\frac{1}{6} = 16.67\%$ ), the accuracy score achieved from these results were less than satisfactory.

From the distribution of class samples of the dataset used for the model training (Table 3.1), there is a severe imbalance in sample size between the classes. From the 6

classes, the majority class (Robust) has 343 samples, whereas the minority class (Frail) has only 7 samples. When conducting the model training, the imbalance of class distribution causes the classification to bias towards the majority samples. With an extremely small number of samples for Frail, the model does not have sufficient examples to learn the characteristics of this class. This scenario is reflected in the confusion matrices produced from the algorithms for the multiclass classification, whereby the model does not predict any samples as Frail. The severe class imbalance that exists in the dataset has contributed to the poor model accuracy for this classification task.

To analyse the classification power of the features used in the dataset, a histogram plot is produced for a select number of features in the dataset. The histogram plot presents the distribution of the values for each feature according to the class membership. The histogram plots in Figure 4.1.

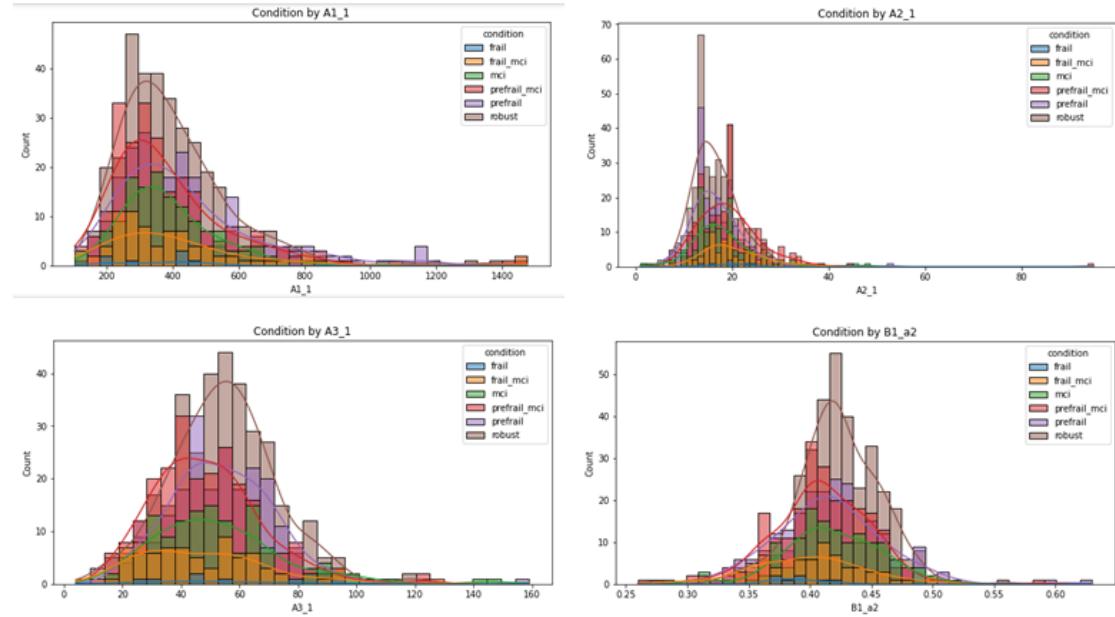


Figure 4.1: Histogram Plots of Feature Value Distribution by Class Membership

From a mathematical and statistical standpoint, the classification and prediction of data samples based on a feature is greatly contributed by the behaviour of feature values for each class. In the case of the dataset used, it was found that most feature values for all the classes overlapped greatly with each other. The severe overlapping between feature values for each class made it difficult for the classification models to accurately predict the class membership of the data samples. As a result, the accuracy of the models were poor.

## 4.2 Binary Classification of Robust and Non-Robust

Table 4.3 and Table 4.4 show the performance evaluation for the binary classification of Robust and Non-Robust for both holdout method and k-fold cross validation. The AUC and ROC Curve plots for each algorithm are shown in Figure 4.2.

Table 4.3: Holdout Method Performance Evaluation (Robust and Non-Robust)

Algorithms	Accuracy	Precision	Recall	F-Measure
Logistic Regression (LOG)	66.5%	67%	55%	51%
Linear Discriminant Analysis (LDA)	65.5%	61%	58%	57%
K-Nearest Neighbor (KNN)	59.9%	55%	54%	54%
Classification and Regression Tree (CART)	55.9%	52%	52%	52%
Gaussian Naive Bayes (GNB)	51.0%	67%	55%	51%
Support Vector Machines (SVM)	63.8%	32%	50%	39%
Random Forest Classifier (RFC)	64.5%	60%	53%	48%

Table 4.4: 5-fold Cross Validation Performance Evaluation (Robust and Non-Robust)

Algorithms	Accuracy	Precision	Recall	F-Measure
Logistic Regression (LOG)	65.7% (std. dev. 0.04)	61.2%	53.4%	49.8%
Linear Discriminant Analysis (LDA)	65.5% (std. dev. 0.05)	61.0%	56.4%	55.6%
K-Nearest Neighbor (KNN)	60.3% (std. dev. 0.04)	54.4%	54.0%	54.0%
Classification and Regression Tree (CART)	57.6% (std. dev. 0.05)	53.6%	53.4%	53.4%
Gaussian Naive Bayes (GNB)	49.4% (std. dev. 0.06)	57.8%	57.2%	48.8%
Support Vector Machines (SVM)	64.9% (std. dev. 0.03)	35.0%	49.4%	40.4%
Random Forest Classifier (RFC)	64.6% (std. dev. 0.02)	60.6%	53.6%	49.8%

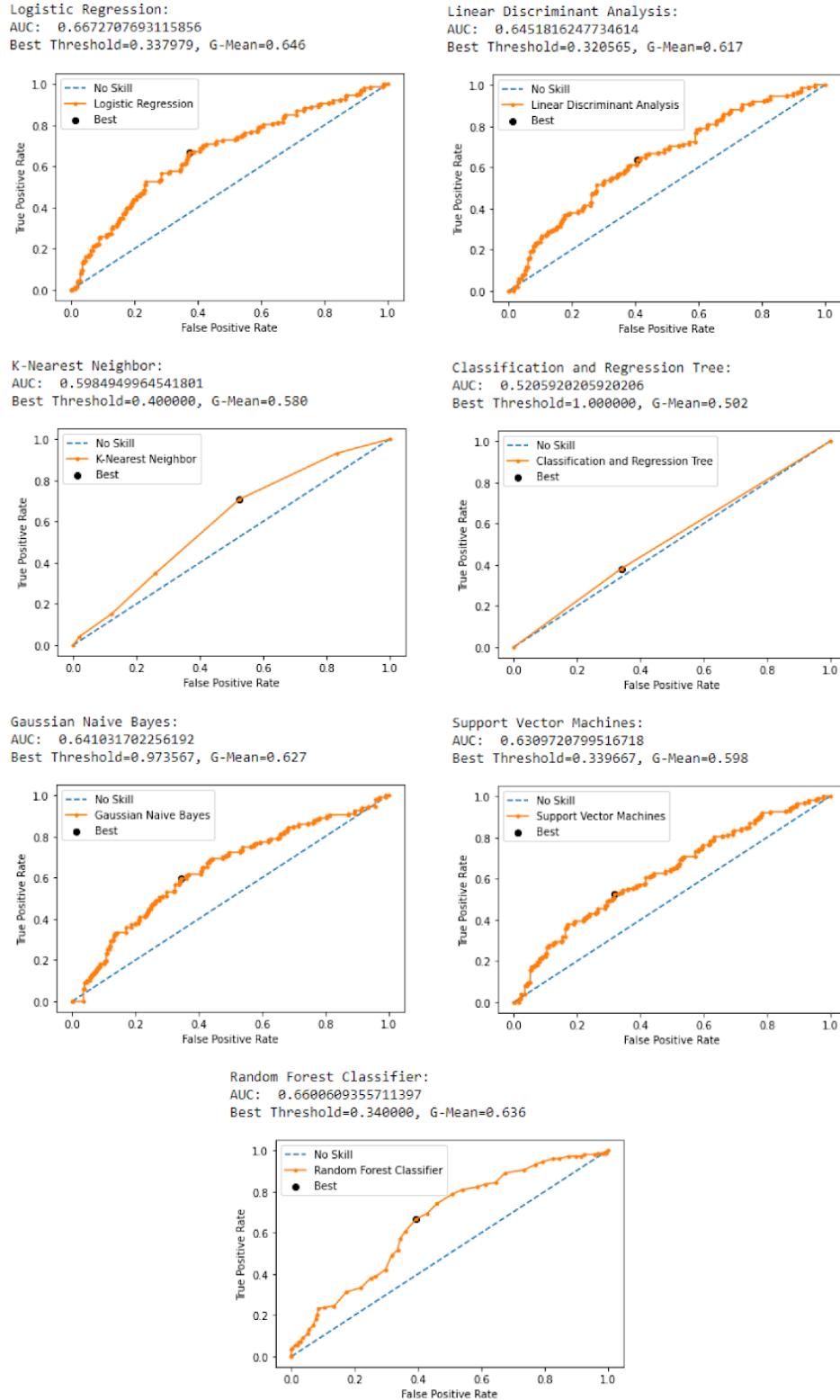


Figure 4.2: AUC and ROC Curves (Robust and Non-Robust)

The binary classification of Robust and Non-Robust as well as Robust and Frail with MCI are attempts at overcoming the class imbalance problem faced in the multiclass classification. By merging the classes aside from Robust together into a single Non-Robust class, the number of samples for each class became 343 and 672 (almost a 1:2 distribution). To further balance the class distribution, the majority class was undersampled to achieve a 1:1 distribution.

For the binary classification for Robust and Non-Robust, LOG had the highest holdout method accuracy of 66.5% and highest cross validation accuracy at 65.7%. The accuracy scores were shown to increase significantly compared to the multiclass classification. The model accuracy was only slightly higher than the probability of classifying into 2 classes ( $\frac{1}{2} = 50.0\%$ ).

While the grouping of classes to allow binary classification produces a better model accuracy, the classification of patients into Robust and Non-Robust does not prove to be useful. This classification task merely distinguishes healthy individuals from non-healthy individuals, with no indication of frailty or MCI in the classification. This task also does not clearly achieve the objectives of the project, which is to classify CF in patients.

### **4.3 Binary Classification of Robust and Frail with MCI**

Table 4.5 – 4.10 show the performance evaluation for the binary classification of Robust and Frail with MCI using different class sample sizes of 76, 343, and 100 for both holdout method and k-fold cross validation. The AUC and ROC Curve plots for each algorithm are shown in Figure 4.3 – 4.5.

Table 4.5: Holdout Method Performance Evaluation (Robust and Frail with MCI – 76 Samples)

Algorithms	Accuracy	Precision	Recall	F-Measure
Logistic Regression (LOG)	55.7%	56%	56%	56%
Linear Discriminant Analysis (LDA)	67.2%	67%	67%	67%
K-Nearest Neighbor (KNN)	54.1%	57%	55%	50%
Classification and Regression Tree (CART)	62.3%	62%	62%	62%
Gaussian Naive Bayes (GNB)	70.5%	56%	56%	56%
Support Vector Machines (SVM)	63.9%	64%	64%	64%
Random Forest Classifier (RFC)	62.3%	62%	62%	62%

Table 4.6: 5-fold Cross Validation Performance Evaluation (Robust and Frail with MCI – 76 Samples)

Algorithms	Accuracy	Precision	Recall	F-Measure
Logistic Regression (LOG)	73.0% (std. dev. 0.07)	73.6%	73.2%	73.0%
Linear Discriminant Analysis (LDA)	64.5% (std. dev. 0.04)	64.8%	64.6%	64.4%
K-Nearest Neighbor (KNN)	67.1% (std. dev. 0.05)	70.0%	67.0%	65.8%
Classification and Regression Tree (CART)	54.6% (std. dev. 0.06)	54.2%	53.8%	53.6%
Gaussian Naive Bayes (GNB)	75.0% (std. dev. 0.09)	76.2%	75.2%	74.6%

Support Vector Machines (SVM)	72.4% (std. dev. 0.05)	73.2%	72.6%	72.4%
Random Forest Classifier (RFC)	73.0% (std. dev. 0.10)	71.8%	71.4%	71.0%

Table 4.7: Holdout Method Performance Evaluation (Robust and Frail with MCI – 343 Samples)

Algorithms	Accuracy	Precision	Recall	F-Measure
Logistic Regression (LOG)	77.8%	78%	78%	78%
Linear Discriminant Analysis (LDA)	82.2%	83%	82%	82%
K-Nearest Neighbor (KNN)	75.6%	80%	75%	74%
Classification and Regression Tree (CART)	77.1%	77%	77%	77%
Gaussian Naive Bayes (GNB)	76.4%	78%	78%	78%
Support Vector Machines (SVM)	79.3%	80%	79%	79%
Random Forest Classifier (RFC)	90.2%	90%	90%	90%

Table 4.8: 5-fold Cross Validation Performance Evaluation (Robust and Frail with MCI – 343 Samples)

Algorithms	Accuracy	Precision	Recall	F-Measure
Logistic Regression (LOG)	80.3% (std. dev. 0.03)	80.6%	80.4%	80.0%
Linear Discriminant Analysis (LDA)	79.3% (std. dev. 0.03)	79.8%	79.2%	79.0%

K-Nearest Neighbor (KNN)	77.3% (std. dev. 0.03)	82.8%	77.2%	76.2%
Classification and Regression Tree (CART)	80.5% (std. dev. 0.04)	81.4%	80.8%	80.8%
Gaussian Naive Bayes (GNB)	73.9% (std. dev. 0.03)	74.4%	74.2%	74.0%
Support Vector Machines (SVM)	79.4% (std. dev. 0.04)	80.8%	79.6%	79.2%
Random Forest Classifier (RFC)	92.0% (std. dev. 0.04)	93.0%	92.6%	92.6%

Table 4.9: Holdout Method Performance Evaluation (Robust and Frail with MCI – 100 Samples)

Algorithms	Accuracy	Precision	Recall	F-Measure
Logistic Regression (LOG)	80.0%	80%	80%	80%
Linear Discriminant Analysis (LDA)	73.8%	74%	74%	74%
K-Nearest Neighbor (KNN)	72.5%	74%	72%	72%
Classification and Regression Tree (CART)	76.2%	79%	77%	76%
Gaussian Naive Bayes (GNB)	75.0%	80%	80%	80%
Support Vector Machines (SVM)	80.0%	80%	80%	80%
Random Forest Classifier (RFC)	82.5%	83%	83%	82%

Table 4.10: 5-fold Cross Validation Performance Evaluation (Robust and Frail with MCI – 100 Samples)

Algorithms	Accuracy	Precision	Recall	F-Measure
Logistic Regression (LOG)	78.0% (std. dev. 0.07)	78.6%	78.0%	77.6%
Linear Discriminant Analysis (LDA)	77.5% (std. dev. 0.09)	78.6%	77.4%	77.2%
K-Nearest Neighbor (KNN)	73.0% (std. dev. 0.10)	74.8%	72.8%	72.6%
Classification and Regression Tree (CART)	78.5% (std. dev. 0.09)	75.6%	75.6%	75.4%
Gaussian Naive Bayes (GNB)	74.5% (std. dev. 0.05)	77.6%	74.2%	73.4%
Support Vector Machines (SVM)	79.5% (std. dev. 0.08)	79.8%	79.4%	79.4%
Random Forest Classifier (RFC)	84.0% (std. dev. 0.10)	85.2%	85.0%	84.8%

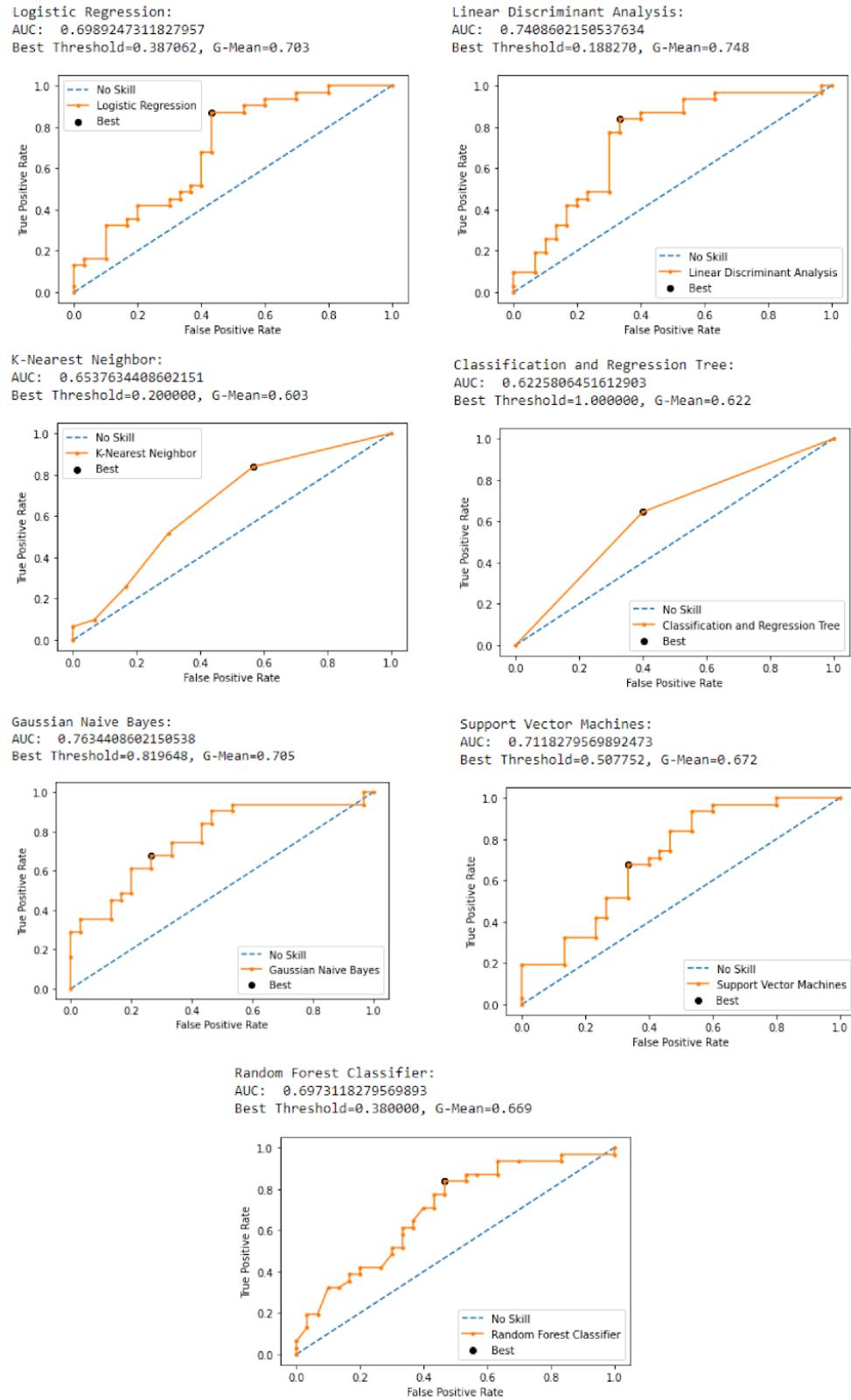


Figure 4.3: AUC and ROC Curves (Robust and Frail with MCI – 76 samples)

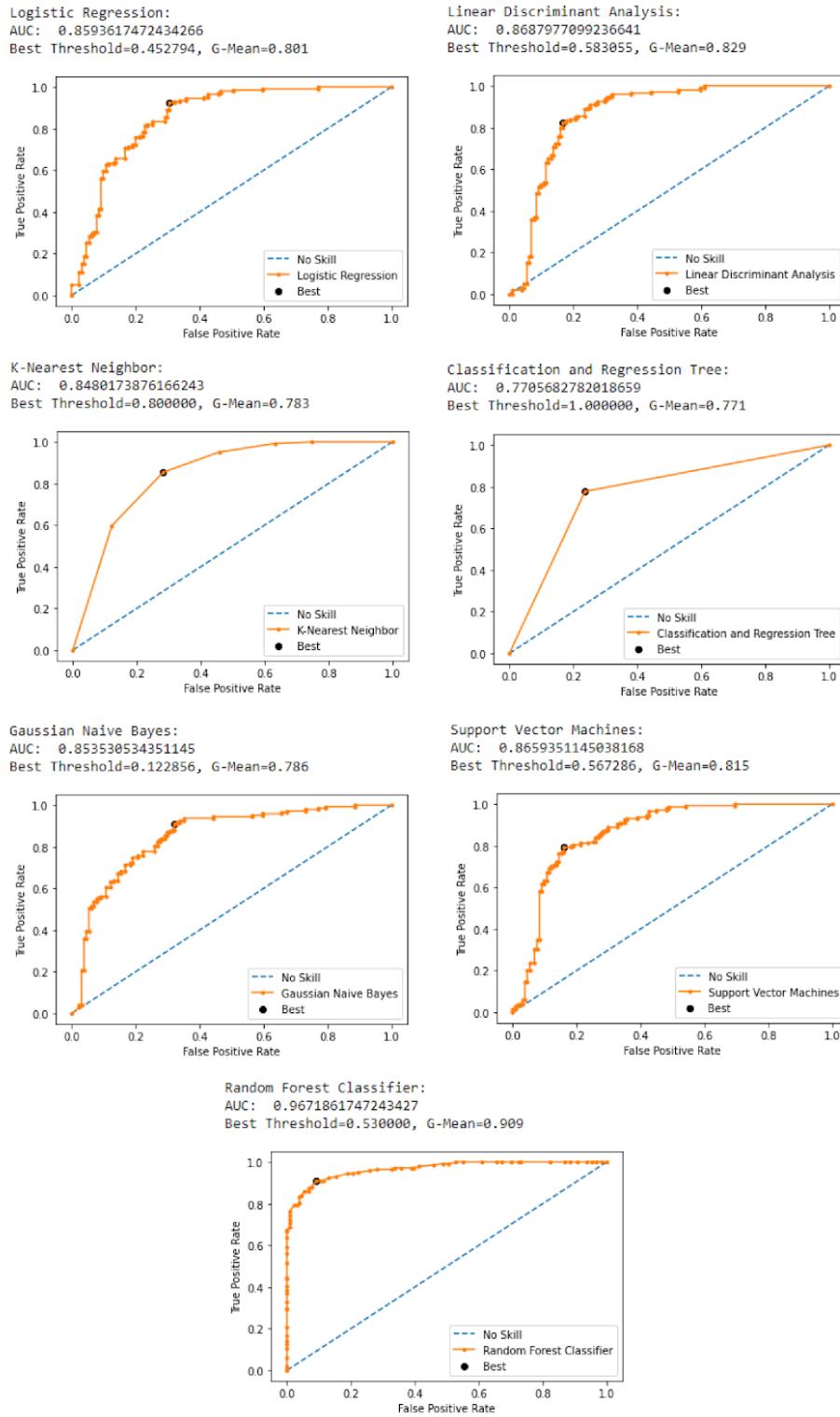


Figure 4.4: AUC and ROC Curves (Robust and Frail with MCI – 343 samples)

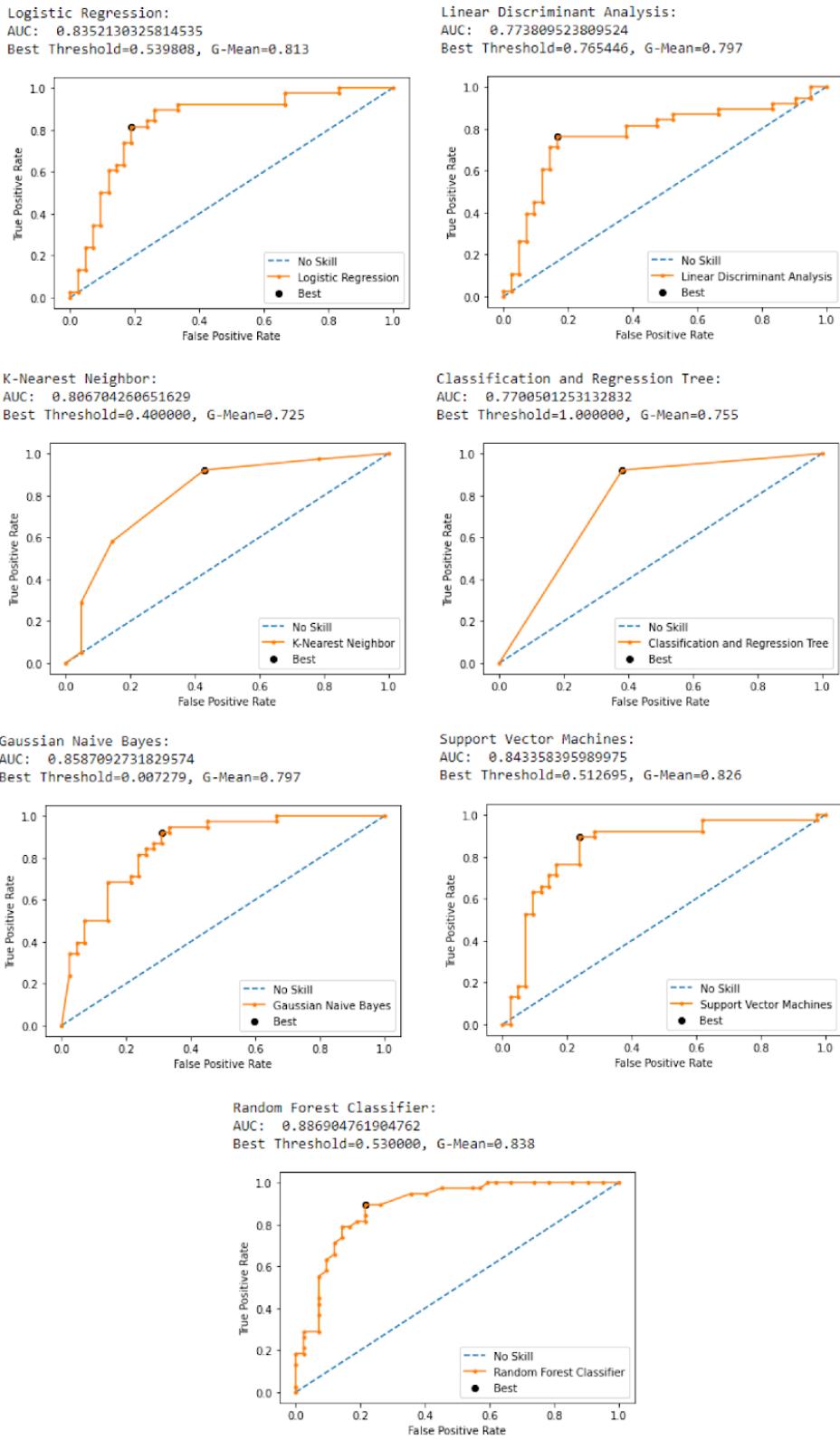


Figure 4.5: AUC and ROC Curves (Robust and Frail with MCI – 100 samples)

The classification of Robust and Frail with MCI aims to distinguish between the extreme ends of the disease spectrum, whereby the healthiest individuals are compared with the least healthiest individuals (in terms of CF). The first class sample distribution undersamples Robust from 343 samples to 76 samples, in order to achieve a 1:1 class distribution. Using this class distribution, GNB showed the highest holdout method accuracy of 70.5%, as well as highest cross validation accuracy of 75.0%.

The model accuracy is further improved compared to the classification of Robust and Non-Robust. Compared to the previous classification task, the Robust and Frail with MCI classification is also better at achieving the project objectives. Patients with both frailty and MCI can be distinguished from healthy individuals through this classification task. While the model performance and classification task are promising, there is concern with regards to the sample size used for model training. The better scores in model performance evaluation could possibly be attributed to a small test dataset size, rather than actual good model performance. Having 76 samples for each class, the total dataset used for both model training and testing (in the case of the holdout method) adds up to 152 samples. While there is no clear rule when it comes to the minimum number of samples needed to accomplish classification using machine learning, it may be useful to explore opportunities in increasing the dataset size to produce a better classification model.

SMOTE was used to oversample the dataset, in order to balance the class distribution and increase the dataset size. As discussed in the literature review, there is no clear indication of the maximum allowed percentage of the dataset that can be SMOTE-ed. A quick analysis was done to understand the relationship between varying SMOTE amounts and the model accuracy and AUC. Model training and performance evaluation was done iteratively by varying the undersampling of the majority class (Robust). After undersampling, the minority class (Frail with MCI) is SMOTE-ed to match the class size of the majority class. Figure 4.6 and Figure 4.7 describe the changes in model accuracy and AUC with majority class size.

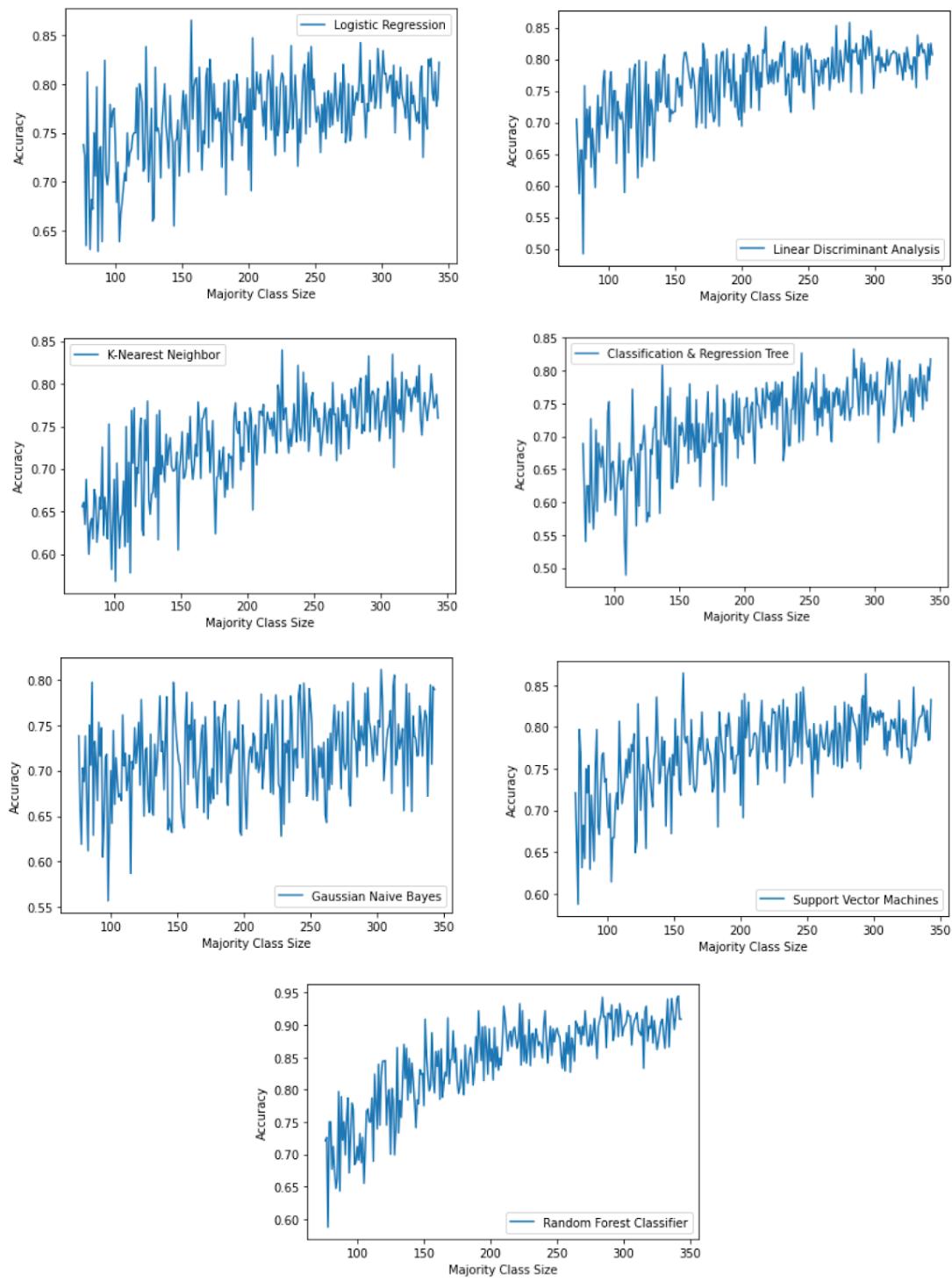


Figure 4.6: Plot of Accuracy Scores against Majority Class Size

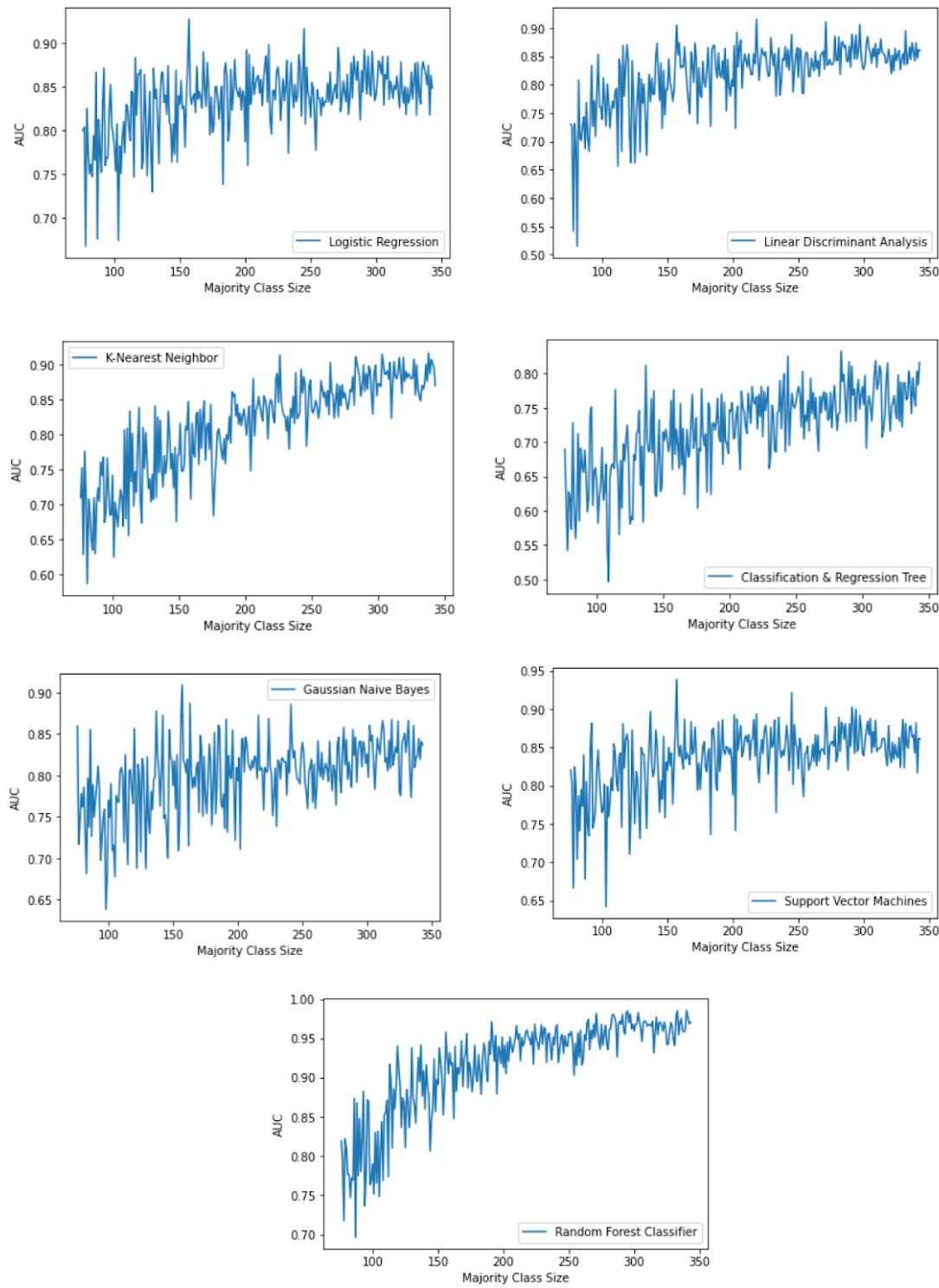


Figure 4.7: Plot of AUC against Majority Class Size

The fluctuating behaviour of both model accuracy and AUC is associated with the stochastic nature of the classification algorithms or evaluation procedures. However, it is observed that the fluctuation reduces (graph becomes more stable) as the majority class size increases. A higher model accuracy and AUC tends to indicate a better model performance (this is not entirely true). As the majority class size increases, both the accuracy and AUC increase. This is due to an increase in the number of samples being SMOTE-ed. Since SMOTE conducts oversampling by selecting and drawing samples from existing samples of the minority class, the SMOTE-ed samples are much more predictable for the classification models. This predictability of the minority class contributes to the increase of model accuracy and AUC.

For the second distribution (343 samples, whereby the minority class of 76 samples are SMOTE-ed entirely), the highest holdout method accuracy and highest cross validation accuracy came from the RFC model, with accuracy scores of 90.2% and 92.0% respectively. The model accuracy was the highest among all the classification tasks and class distributions. As described previously, the high performance in evaluation scores are associated with the predictability of SMOTE-ed samples. As such, while the model training and testing does provide good results, it is not a good option to apply the models from this approach for the disease classification.

Since SMOTE-ing too many samples creates an unreliable model, there is a need to only SMOTE a small percentage of the dataset, while maintaining a balanced class distribution. The third distribution (100 samples) first undersamples the majority class to achieve a 50:50 distribution, then SMOTEs both classes to obtain a sufficient number of samples. For this distribution, RFC achieved the highest holdout method accuracy of 82.5% and the highest cross validation accuracy of 84.0%. The model accuracy is not as high as the second distribution (343 samples), due to less SMOTE-ed samples and subsequently less predictability by the model. However, by conducting SMOTE for a small number of samples, there is now a sufficient number of samples for the dataset (200 samples), while keeping the reliability of the model in control. This is reflected in

the confusion matrices for the classification as shown in Figure 4.8, whereby the number of true positive and true negative predictions are higher than the false positive and false negative predictions.

```
Performance Metrics for RFC:  
0.825  
[[32 10]  
 [ 4 34]]  


|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.89      | 0.76   | 0.82     | 42      |
| 1            | 0.77      | 0.89   | 0.83     | 38      |
| accuracy     |           |        | 0.82     | 80      |
| macro avg    | 0.83      | 0.83   | 0.82     | 80      |
| weighted avg | 0.83      | 0.82   | 0.82     | 80      |


```

Figure 4.8: Confusion Matrix and Performance Evaluation for RFC (Robust and Frail with MCI) – 100 samples

#### 4.4 Significant Blood Biomarkers in CF

After the models were trained and tested, the features in the dataset were recursively eliminated using RFE. As the features were reduced, the models were evaluated in terms of accuracy. From the RFE conducted, the top 10 features for the LOG, LDA, CART, SVM, and RFC algorithms were obtained. These features are provided in Tables 4.11 – 4.15.

Table 4.11: Top 10 Ranked Features – 6 Classes

Logistic Regression	Linear Discriminant Analysis
1. Potassium (mmol/L) 1. Serum Folate (nmol/L) 1. Serum Homocysteine ( $\mu$ mol/L) 1. Lymphocytes (/L) 1. Neutrophils (/L) 1. RBC (/L) 1. PCV (L/L) 1. Chloride (mmol/L) 1. Free Tri-iodothyronine (FT3) (pmol/L) 1. MCHC (g/L) 2. Sodium (mmol/L) 3. Haemoglobin (g/L) 4. White Cell Count (/L) 5. Monocytes (/L) 6. Platelets (/L) 7. Glucose (mmol/L) 8. Total Protein (g/L) 9. Vitamin B12 (pmol/L) 10. Total Cholesterol/HDL Ratio	1. LDL Cholesterol (mmol/L) 1. Thyroid Stimulating Hormone (mIU/L) 1. Total Cholesterol (mmol/L) 1. Creatinine (umol/L) 1. Urea (mmol/L) 1. RBC (/L) 1. PCV (L/L) 1. MCV (fL) 1. AST 1. ALT 2. Free Tri-iodothyronine (FT3) (pmol/L) 3. White Cell Count (/L) 4. RDW (%) 5. MCH (pg) 6. MCHC (g/L) 7. Total Bilirubin ( $\mu$ mol/L) 8. Serum Homocysteine ( $\mu$ mol/L) 9. Neutrophils (/L) 10. Albumin/Globulin ratio
Classification & Regression Tree	Support Vector Machines
1. Urea (mmol/L) 1. Serum Folate (nmol/L) 1. Serum Homocysteine ( $\mu$ mol/L) 1. 25-hydroxy Vitamin D (nmol/L) 1. RBC (/L) 1. Thyroid Stimulating Hormone (mIU/L) 1. Glucose (mmol/L) 1. Triglyceride (mmol/L) 1. LDL Cholesterol (mmol/L) 1. Neutrophils (/L) 2. Haemoglobin (g/L) 3. Calcium (mmol/L) 4. Globulin (g/L) 5. Potassium (mmol/L) 6. Uric Acid (mmol/L) 7. Monocytes (/L) 8. Corrected Calcium (mmol/L) 9. Platelets (/L) 10. Total Bilirubin ( $\mu$ mol/L)	1. Chloride (mmol/L) 1. Serum Folate (nmol/L) 1. Serum Homocysteine ( $\mu$ mol/L) 1. Haemoglobin (g/L) 1. RBC (/L) 1. PCV (L/L) 1. Potassium (mmol/L) 1. Lymphocytes (/L) 1. Neutrophils (/L) 1. White Cell Count (/L) 2. HDL Cholesterol (mmol/L) 3. Corrected Calcium (mmol/L) 4. Vitamin B12 (pmol/L) 5. MCHC (g/L) 6. Total Cholesterol/HDL Ratio 7. Platelets (/L) 8. Triglyceride (mmol/L) 9. Total Protein (g/L) 10. Monocytes (/L)
Random Forest Classifier	
1. Vitamin B12 (pmol/L) 1. Serum Folate (nmol/L) 1. 25-hydroxy Vitamin D (nmol/L) 1. Haemoglobin (g/L) 1. RBC (/L) 1. Thyroid Stimulating Hormone (mIU/L) 1. LDL Cholesterol (mmol/L) 1. HDL Cholesterol (mmol/L) 1. Neutrophils (/L) 1. Glucose (mmol/L) 2. Triglyceride (mmol/L) 3. White Cell Count (/L) 4. Creatinine (umol/L) 5. Serum Homocysteine ( $\mu$ mol/L) 6. Platelets (/L) 7. GGT 8. Urea (mmol/L) 9. C-Reactive Protein 10. Alkaline Phosphatase (U/L)	

Table 4.12: Top 10 Ranked Features – Robust & Non-Robust

Logistic Regression	Linear Discriminant Analysis
1. Platelets (/L) 1. Serum Homocysteine ( $\mu\text{mol/L}$ ) 1. 25-hydroxy Vitamin D (nmol/L) 1. Monocytes (/L) 1. RBC (/L) 1. PCV (L/L) 1. Free Tri-iodothyronine (FT3) (pmol/L) 1. Chloride (mmol/L) 1. Lymphocytes (/L) 1. Neutrophils (/L) 2. Sodium (mmol/L) 3. pH 4. HDL Cholesterol (mmol/L) 5. Potassium (mmol/L) 6. RDW (%) 7. Corrected Calcium (mmol/L) 8. Serum Folate (nmol/L) 9. Haemoglobin (g/L) 10. Vitamin B12 (pmol/L)	1. Calcium (mmol/L) 1. Corrected Calcium (mmol/L) 1. Serum Homocysteine ( $\mu\text{mol/L}$ ) 1. Total Protein (g/L) 1. Albumin (g/L) 1. RBC (/L) 1. Monocytes (/L) 1. Neutrophils (/L) 1. Globulin (g/L) 1. White Cell Count (/L) 2. Creatinine (umol/L) 3. Platelets (/L) 4. Lymphocytes (/L) 5. Chloride (mmol/L) 6. Thyroid Stimulating Hormone (mIU/L) 7. MCH (pg) 8. PCV (L/L) 9. MCHC (g/L) 10. Albumin/Globulin ratio
Classification & Regression Tree	Support Vector Machines
1. Platelets (/L) 1. Serum Folate (nmol/L) 1. Serum Homocysteine ( $\mu\text{mol/L}$ ) 1. 25-hydroxy Vitamin D (nmol/L) 1. Corrected Calcium (mmol/L) 1. Thyroid Stimulating Hormone (mIU/L) 1. Creatinine (umol/L) 1. Total Cholesterol (mmol/L) 1. C-Reactive Protein 1. Neutrophils (/L) 2. ALT 3. HDL Cholesterol (mmol/L) 4. RBC (/L) 5. HbA1c 6. Free Tri-iodothyronine (FT3) (pmol/L) 7. LDL Cholesterol (mmol/L) 8. Alkaline Phosphatase (U/L) 9. Glucose (mmol/L) 10. Calcium (mmol/L)	1. White Cell Count (/L) 1. Serum Homocysteine ( $\mu\text{mol/L}$ ) 1. RBC (/L) 1. Creatinine (umol/L) 1. Globulin (g/L) 1. Total Protein (g/L) 1. Platelets (/L) 1. Monocytes (/L) 1. Neutrophils (/L) 1. Lymphocytes (/L) 2. Chloride (mmol/L) 3. HDL Cholesterol (mmol/L) 4. Thyroid Stimulating Hormone (mIU/L) 5. 25-hydroxy Vitamin D (nmol/L) 6. Free Tri-iodothyronine (FT3) (pmol/L) 7. Uric Acid (mmol/L) 8. RDW (%) 9. Total Cholesterol (mmol/L) 10. pH
Random Forest Classifier	
1. Vitamin B12 (pmol/L) 1. Serum Folate (nmol/L) 1. Thyroid Stimulating Hormone (mIU/L) 1. 25-hydroxy Vitamin D (nmol/L) 1. Haemoglobin (g/L) 1. RBC (/L) 1. LDL Cholesterol (mmol/L) 1. Platelets (/L) 1. Alkaline Phosphatase (U/L) 1. Neutrophils (/L) 2. Glucose (mmol/L) 3. Triglyceride (mmol/L) 4. Creatinine (umol/L) 5. White Cell Count (/L) 6. HDL Cholesterol (mmol/L) 7. GGT 8. Total Cholesterol (mmol/L) 9. ALT 10. RDW (%)	

Table 4.13: Top 10 Ranked Features – Robust and Frail with MCI (76 samples)

Logistic Regression	Linear Discriminant Analysis
1. Sodium (mmol/L) 1. Total Bilirubin ( $\mu\text{mol/L}$ ) 1. Monocytes (/L) 1. ALT 1. Thyroid Stimulating Hormone (mIU/L) 1. PCV (L/L) 1. Lymphocytes (/L) 1. Serum Folate (nmol/L) 1. Serum Homocysteine ( $\mu\text{mol/L}$ ) 1. RBC (/L) 2. Glucose (mmol/L) 3. Haemoglobin (g/L) 4. 25-hydroxy Vitamin D (nmol/L) 5. Chloride (mmol/L) 6. Phosphate (mmol/L) 7. Albumin/Globulin ratio 8. Creatinine ( $\mu\text{mol/L}$ ) 9. MCV (fL) 10. Neutrophils (/L)	1. Thyroid Stimulating Hormone (mIU/L) 1. RBC (/L) 1. Albumin (g/L) 1. MCV (fL) 1. MCH (pg) 1. Globulin (g/L) 1. C-Reactive Protein 1. Albumin/Globulin ratio 1. Total Bilirubin ( $\mu\text{mol/L}$ ) 1. Platelets (/L) 2. Lymphocytes (/L) 3. Triglyceride (mmol/L) 4. Serum Homocysteine ( $\mu\text{mol/L}$ ) 5. Free Tri-iodothyronine (FT3) (pmol/L) 6. AST 7. Chloride (mmol/L) 8. Vitamin B12 (pmol/L) 9. Total Protein (g/L) 10. Glucose (mmol/L)
Classification & Regression Tree	Support Vector Machines
1. Total Bilirubin ( $\mu\text{mol/L}$ ) 1. Serum Folate (nmol/L) 1. AST 1. C-Reactive Protein 1. Haemoglobin (g/L) 1. Albumin/Globulin ratio 1. PCV (L/L) 1. Globulin (g/L) 1. Albumin (g/L) 1. White Cell Count (/L) 2. Phosphate (mmol/L) 3. MCV (fL) 4. Total Protein (g/L) 5. MCH (pg) 6. RBC (/L) 7. 25-hydroxy Vitamin D (nmol/L) 8. Serum Homocysteine ( $\mu\text{mol/L}$ ) 9. MCHC (g/L) 10. RDW (%)	1. ALT 1. Monocytes (/L) 1. Total Bilirubin ( $\mu\text{mol/L}$ ) 1. Glucose (mmol/L) 1. PCV (L/L) 1. Chloride (mmol/L) 1. Thyroid Stimulating Hormone (mIU/L) 1. Serum Homocysteine ( $\mu\text{mol/L}$ ) 1. Serum Folate (nmol/L) 1. RBC (/L) 2. Lymphocytes (/L) 3. Sodium (mmol/L) 4. Corrected Calcium (mmol/L) 5. Albumin/Globulin ratio 6. MCHC (g/L) 7. Albumin (g/L) 8. MCV (fL) 9. Haemoglobin (g/L) 10. Vitamin B12 (pmol/L)
Random Forest Classifier	
1. Free Tri-iodothyronine (FT3) (pmol/L) 1. Serum Folate (nmol/L) 1. Serum Homocysteine ( $\mu\text{mol/L}$ ) 1. 25-hydroxy Vitamin D (nmol/L) 1. Haemoglobin (g/L) 1. RBC (/L) 1. PCV (L/L) 1. Phosphate (mmol/L) 1. ALT 1. White Cell Count (/L) 2. Lymphocytes (/L) 3. Urea (mmol/L) 4. Glucose (mmol/L) 5. Neutrophils (/L) 6. Thyroid Stimulating Hormone (mIU/L) 7. MCV (fL) 8. Uric Acid (mmol/L) 9. Monocytes (/L) 10. Total Bilirubin ( $\mu\text{mol/L}$ )	

Table 4.14: Top 10 Ranked Features – Robust and Frail with MCI (343 samples)

Logistic Regression	Linear Discriminant Analysis
<b>1.</b> Sodium (mmol/L) <b>1.</b> Serum Folate (nmol/L) <b>1.</b> Serum Homocysteine ( $\mu$ mol/L) <b>1.</b> 25-hydroxy Vitamin D (nmol/L) <b>1.</b> C-Reactive Protein <b>1.</b> RBC (/L) <b>1.</b> PCV (L/L) <b>1.</b> ALT <b>1.</b> Total Bilirubin ( $\mu$ mol/L) <b>1.</b> Potassium (mmol/L) <b>2.</b> Monocytes (/L) <b>3.</b> HbA1c <b>4.</b> Creatinine (umol/L) <b>5.</b> S.G. <b>6.</b> Chloride (mmol/L) <b>7.</b> Total Protein (g/L) <b>8.</b> pH <b>9.</b> Neutrophils (/L) <b>10.</b> Vitamin B12 (pmol/L)	<b>1.</b> MCH (pg) <b>1.</b> Globulin (g/L) <b>1.</b> MCHC (g/L) <b>1.</b> Albumin (g/L) <b>1.</b> PCV (L/L) <b>1.</b> Total Protein (g/L) <b>1.</b> ALT <b>1.</b> C-Reactive Protein <b>1.</b> Serum Homocysteine ( $\mu$ mol/L) <b>1.</b> RBC (/L) <b>2.</b> AST <b>3.</b> GGT <b>4.</b> Total Bilirubin ( $\mu$ mol/L) <b>5.</b> Sodium (mmol/L) <b>6.</b> Serum Folate (nmol/L) <b>7.</b> Alkaline Phosphatase (U/L) <b>8.</b> Neutrophils (/L) <b>9.</b> White Cell Count (/L) <b>10.</b> HbA1c
Classification & Regression Tree	Support Vector Machines
<b>1.</b> C-Reactive Protein <b>1.</b> Serum Homocysteine ( $\mu$ mol/L) <b>1.</b> 25-hydroxy Vitamin D (nmol/L) <b>1.</b> Triglyceride (mmol/L) <b>1.</b> Potassium (mmol/L) <b>1.</b> PCV (L/L) <b>1.</b> Free Tri-iodothyronine (FT3) (pmol/L) <b>1.</b> White Cell Count (/L) <b>1.</b> Glucose (mmol/L) <b>1.</b> Creatinine (umol/L) <b>2.</b> Albumin (g/L) <b>3.</b> MCHC (g/L) <b>4.</b> Calcium (mmol/L) <b>5.</b> ALT <b>6.</b> RDW (%) <b>7.</b> RBC (/L) <b>8.</b> Platelets (/L) <b>9.</b> Vitamin B12 (pmol/L) <b>10.</b> Haemoglobin (g/L)	<b>1.</b> Sodium (mmol/L) <b>1.</b> Monocytes (/L) <b>1.</b> Total Bilirubin ( $\mu$ mol/L) <b>1.</b> GGT <b>1.</b> C-Reactive Protein <b>1.</b> PCV (L/L) <b>1.</b> ALT <b>1.</b> Serum Homocysteine ( $\mu$ mol/L) <b>1.</b> Serum Folate (nmol/L) <b>1.</b> RBC (/L) <b>2.</b> Total Protein (g/L) <b>3.</b> Potassium (mmol/L) <b>4.</b> Creatinine (umol/L) <b>5.</b> S.G. <b>6.</b> Chloride (mmol/L) <b>7.</b> pH <b>8.</b> Neutrophils (/L) <b>9.</b> Globulin (g/L) <b>10.</b> 25-hydroxy Vitamin D (nmol/L)
Random Forest Classifier	
<b>1.</b> Serum Folate (nmol/L) <b>1.</b> Serum Homocysteine ( $\mu$ mol/L) <b>1.</b> 25-hydroxy Vitamin D (nmol/L) <b>1.</b> Neutrophils (/L) <b>1.</b> RBC (/L) <b>1.</b> PCV (L/L) <b>1.</b> C-Reactive Protein <b>1.</b> Free Tri-iodothyronine (FT3) (pmol/L) <b>1.</b> White Cell Count (/L) <b>1.</b> ALT <b>2.</b> HbA1c <b>3.</b> MCH (pg) <b>4.</b> Triglyceride (mmol/L) <b>5.</b> S.G. <b>6.</b> Potassium (mmol/L) <b>7.</b> MCHC (g/L) <b>8.</b> Calcium (mmol/L) <b>9.</b> Creatinine (umol/L) <b>10.</b> Sodium (mmol/L)	

Table 4.15: Top 10 Ranked Features – Robust and Frail with MCI (100 samples)

Logistic Regression	Linear Discriminant Analysis
<b>1.</b> Sodium (mmol/L) <b>1.</b> Serum Folate (nmol/L) <b>1.</b> Serum Homocysteine ( $\mu$ mol/L) <b>1.</b> pH <b>1.</b> ALT <b>1.</b> RBC (/L) <b>1.</b> PCV (L/L) <b>1.</b> Creatinine (umol/L) <b>1.</b> Chloride (mmol/L) <b>1.</b> Glucose (mmol/L) <b>2.</b> Phosphate (mmol/L) <b>3.</b> Vitamin B12 (pmol/L) <b>4.</b> Urea (mmol/L) <b>5.</b> HbA1c <b>6.</b> LDL Cholesterol (mmol/L) <b>7.</b> MCV (fL) <b>8.</b> White Cell Count (/L) <b>9.</b> Haemoglobin (g/L) <b>10.</b> C-Reactive Protein	<b>1.</b> C-Reactive Protein <b>1.</b> Total Protein (g/L) <b>1.</b> Albumin (g/L) <b>1.</b> Total Cholesterol/HDL Ratio <b>1.</b> LDL Cholesterol (mmol/L) <b>1.</b> Globulin (g/L) <b>1.</b> Triglyceride (mmol/L) <b>1.</b> Total Cholesterol (mmol/L) <b>1.</b> Albumin/Globulin ratio <b>1.</b> Glucose (mmol/L) <b>2.</b> PCV (L/L) <b>3.</b> GGT <b>4.</b> AST <b>5.</b> Chloride (mmol/L) <b>6.</b> Serum Folate (nmol/L) <b>7.</b> Vitamin B12 (pmol/L) <b>8.</b> Calcium (mmol/L) <b>9.</b> Corrected Calcium (mmol/L) <b>10.</b> Platelets (/L)
Classification & Regression Tree	Support Vector Machines
<b>1.</b> LDL Cholesterol (mmol/L) <b>1.</b> Potassium (mmol/L) <b>1.</b> pH <b>1.</b> Triglyceride (mmol/L) <b>1.</b> Glucose (mmol/L) <b>1.</b> White Cell Count (/L) <b>1.</b> C-Reactive Protein <b>1.</b> Serum Folate (nmol/L) <b>1.</b> PCV (L/L) <b>1.</b> 25-hydroxy Vitamin D (nmol/L) <b>2.</b> Chloride (mmol/L) <b>3.</b> Urea (mmol/L) <b>4.</b> MCH (pg) <b>5.</b> MCV (fL) <b>6.</b> RBC (/L) <b>7.</b> Haemoglobin (g/L) <b>8.</b> Creatinine (umol/L) <b>9.</b> ALT <b>10.</b> Serum Homocysteine ( $\mu$ mol/L)	<b>1.</b> Sodium (mmol/L) <b>1.</b> Serum Folate (nmol/L) <b>1.</b> Serum Homocysteine ( $\mu$ mol/L) <b>1.</b> pH <b>1.</b> RBC (/L) <b>1.</b> PCV (L/L) <b>1.</b> ALT <b>1.</b> Creatinine (umol/L) <b>1.</b> Chloride (mmol/L) <b>1.</b> Glucose (mmol/L) <b>2.</b> C-Reactive Protein <b>3.</b> Albumin/Globulin ratio <b>4.</b> HbA1c <b>5.</b> Vitamin B12 (pmol/L) <b>6.</b> Alkaline Phosphatase (U/L) <b>7.</b> White Cell Count (/L) <b>8.</b> Phosphate (mmol/L) <b>9.</b> Potassium (mmol/L) <b>10.</b> 25-hydroxy Vitamin D (nmol/L)
Random Forest Classifier	
<b>1.</b> HbA1c <b>1.</b> Serum Folate (nmol/L) <b>1.</b> 25-hydroxy Vitamin D (nmol/L) <b>1.</b> RBC (/L) <b>1.</b> PCV (L/L) <b>1.</b> pH <b>1.</b> C-Reactive Protein <b>1.</b> Creatinine (umol/L) <b>1.</b> Neutrophils (/L) <b>1.</b> Glucose (mmol/L) <b>2.</b> White Cell Count (/L) <b>3.</b> Haemoglobin (g/L) <b>4.</b> Serum Homocysteine ( $\mu$ mol/L) <b>5.</b> Free Tri-iodothyronine (FT3) (pmol/L) <b>6.</b> LDL Cholesterol (mmol/L) <b>7.</b> ALT <b>8.</b> Vitamin B12 (pmol/L) <b>9.</b> Chloride (mmol/L) <b>10.</b> Urea (mmol/L)	

After the classification models were evaluated, the features in the dataset used were recursively eliminated to obtain the top features (best performing features). Out of the 46 features in the dataset, the features ranked 1 to 10 are selected and shown as the top features. While the ranking was done from 1 to 10, there were a large number of features that were given rank 1. This scenario could be caused by the equal strength of these features in predicting CF or the generally poor classification strength of the entire dataset as reflected previously.

From the list of features selected, a high repetition of features selected by different classification algorithms was shown. For example, Serum Folate, Serum Homocysteine, and C-Reactive Protein were repeatedly selected as top features by the algorithms. This scenario gives a good indication on the reliability of the feature selection conducted, as well as the confidence in the classification strength of these selected features.

Further analysing the features selected, there is a strong correlation between the obtained features and the literature review conducted on potential biomarkers for the frailty and MCI syndromes. From the literature review, C-Reactive Protein and HbA1c were identified as potential biomarkers and indicators of frailty and MCI. Aside from the literature review, consultation and advice from domain experts (medical researchers) have also indicated Serum Folate and Serum Homocysteine as potential biomarkers. From the table of features selected, C-Reactive Protein, HbA1c, Serum Folate, and Serum Homocysteine have also been identified as the top features obtained from the classification. The strong correlation between potential biomarkers identified from the background study and the top features selected illustrates the ability of the classification models in identifying the correct biomarkers for CF.

## **CHAPTER 5**

### **CONCLUSION AND RECOMMENDATION**

#### **5.1 Conclusion**

With the complexity and difficulty of managing late-stage effects of Cognitive Frailty (CF), a method of predicting the syndrome, along with the contributing biomarkers, are needed to help researchers develop clinical inferences on CF. This study proposes a model using blood-based parameters to classify patients into different levels of CF, while highlighting the most important blood biomarkers relating to CF.

From the study, different models were investigated to classify patients into different levels of CF, based on parameters obtained from blood samples. By recursively eliminating features from the dataset and evaluating the model performance, top features or blood-based parameters for the classification models were obtained. The obtained parameters provide an indication of the potential biomarkers that can be associated with CF.

From the results of the study, the proposed classification models have the clinical significance to allow early prediction of CF in patients, before the onset of the syndrome and its symptoms. Rather than relying on time-consuming and resource-heavy diagnosis tools, it is possible to predict CF using a blood or serum test to obtain the necessary blood-based parameters. Compared to the existing methods for identifying frailty and MCI, the proposed classification models rely on physiological parameters obtained from the patient to verify the disease, which is more reliable than the current assessments. By

obtaining the top features from the classification models, it is possible for clinicians and medical researchers to infer possible blood-based biomarkers that indicate the current or future presence of CF in a patient. These features or blood-based parameters serve as valuable information to enable early medical intervention for CF. Lastly, the convenience of obtaining the needed data, especially in a clinical or laboratory setting, for disease prediction gives the study an advantage over the existing tools. Once the necessary data is obtained, the proposed classification models provide a means of rapid testing of the syndrome, by running the data through the models to generate the disease prediction results.

## 5.2 Recommendation

While the performance evaluation shows the classification models as promising, there is much room for improvement to enable a comprehensive, clinically-ready CF prediction tool. As shown from the analysis conducted on the distribution of class samples, class imbalance is a severe problem faced by this study. Not only did the class imbalance produce a biased model, the class imbalance also caused the models to be unable to predict Frail at all, greatly impacting the model accuracy and performance. Future efforts should ensure that a well-balanced dataset is used, while ensuring a sufficient number of samples and dataset size is used for model training and testing.

The dataset used for the classification originated from a study conducted on a specific demographic of elderly people. While the prevalence of CF is higher among elderly adults, the disease can manifest in individuals from any demographic. Since the patient data used for the model training and testing comes from an urban Malaysian demographic, a comprehensive and inclusive classification model for CF should include data samples from elderly people coming from different demographics and backgrounds. This is to ensure that the produced classification models are not biased to certain parameters associated with a certain demographic (climate of residence, diet, genetic makeup etc).

## REFERENCES

- [1] J. F. Dartigues and H. Amieva, “Cognitive Frailty: Rational and Definition from an (I.A.N.A./I.A.G.G.) International Consensus Group,” *The Journal of Nutrition, Health & Aging*, vol. 18, no. 1, pp. 95–95, 2013.
- [2] “Open Data - Population by age group, sex and ethnic group, 2010 - 2019e, Malaysia,” *Department of Statistics Malaysia Official Portal*. [Online]. Available: [https://www.dosm.gov.my/v1/index.php?r=column3/accordion&menu\\_id=aHhRYUpWS3B4VXlYaVBOeUF0WFpWUT09](https://www.dosm.gov.my/v1/index.php?r=column3/accordion&menu_id=aHhRYUpWS3B4VXlYaVBOeUF0WFpWUT09). [Accessed: 09-Mar-2021].
- [3] H. Shimada, H. Makizako, T. Doi, K. Tsutsumimoto, S. Lee, and T. Suzuki, “P3-371: Impact of Cognitive Frailty on Daily Activities in Older Persons,” *Alzheimers & Dementia*, vol. 12, 2016.
- [4] C. Cano, R. Samper-Ternent, S. A. Snih, K. Markides, and K. J. Ottenbacher, “Frailty and cognitive impairment as predictors of mortality in older Mexican Americans,” *The journal of nutrition, health & aging*, vol. 16, no. 2, pp. 142–147, 2011.
- [5] N. F. M. Rivan, S. Shahar, N. F. Rajab, D. K. A. Singh, N. C. Din, H. Mahadzir, and T. A. T. A. Hamid, “Cognitive frailty among Malaysian older adults: baseline findings from the LRGS TUA cohort study,” *Clinical Interventions in Aging*, vol. 14, pp. 1343–1352, 2019.
- [6] M. Speechley and M. Tinetti, “Falls and Injuries in Frail and Vigorous Community Elderly Persons,” *Journal of the American Geriatrics Society*, vol. 39, no. 1, pp. 46–52, 1991.
- [7] F. Nourhashemi, S. Andrieu, S. Gillette-Guyonnet, B. Vellas, J. L. Albareda, and H. Grandjean, “Instrumental Activities of Daily Living as a Potential Marker of Frailty: A Study of 7364 Community-Dwelling Elderly Women (the EPIDOS Study),” *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 56, no. 7, 2001.
- [8] L. P. Fried et al., “Frailty in older adults: evidence for a phenotype,” *J. Gerontol. A Biol. Sci. Med. Sci.*, vol. 56, no. 3, pp. M146–56, 2001.

- [9] A. B. Mitnitski, A. J. Mogilner, and K. Rockwood, “Accumulation of Deficits as a Proxy Measure of Aging,” *The Scientific World JOURNAL*, vol. 1, pp. 323–336, 2001.
- [10] K. Rockwood, M. Andrew, and A. Mitnitski, “A Comparison of Two Approaches to Measuring Frailty in Elderly People,” *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 62, no. 7, pp. 738–743, 2007.
- [11] J. C. Prichard, *A treatise on insanity and other disorders affecting the mind*. Philadelphia: Haswell, Barrington, and Haswell, 1837.
- [12] V. A. Kral, “Senescent forgetfulness: benign and malignant,” *The Canadian Medical Association Journal*, vol. 86, no.6, pp. 257-260, 1962.
- [13] R. C. Petersen, G. E. Smith, S. C. Waring, R. J. Ivnik, E. G. Tangalos, and E. Kokmen, “Mild cognitive impairment: clinical characterization and outcome,” *Arch. Neurol.*, vol. 56, no. 3, pp. 303–308, 1999.
- [14] I. Arevalo-Rodriguez et al., “Mini-Mental State Examination (MMSE) for the detection of Alzheimer’s disease and other dementias in people with mild cognitive impairment (MCI),” *Cochrane Database Syst. Rev.*, no. 3, p. CD010783, 2015.
- [15] Z. S. Nasreddine et al., “The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment,” *J. Am. Geriatr. Soc.*, vol. 53, no. 4, pp. 695–699, 2005.
- [16] A. Aggarwal and E. Kean, “Comparison of the folstein mini mental state examination (MMSE) to the Montreal cognitive assessment (MoCA) as a cognitive screening tool in an inpatient rehabilitation setting,” *Neurosci. Med.*, vol. 01, no. 02, pp. 39–42, 2010.
- [17] F. Segai-Gidan, “Cognitive Screening Tools,” *Clinician Reviews*, vol. 23, no. 01, 2013
- [18] A. S. Buchman, R. S. Wilson, J. L. Bienias, R. C. Shah, D. A. Evans, and D. A. Bennett, “Change in body mass index and risk of incident Alzheimer disease,” *Neurology*, vol. 65, no. 6, pp. 892–897, 2005.
- [19] P. A. Boyle, A. S. Buchman, R. S. Wilson, S. E. Leurgans, and D. A. Bennett, “Association of Muscle Strength With the Risk of Alzheimer Disease and the

- Rate of Cognitive Decline in Community-Dwelling Older Persons," *Archives of Neurology*, vol. 66, no. 11, 2009.
- [20] R. Camicioli, D. Howieson, B. Oken, G. Sexton, and J. Kaye, "Motor slowing precedes cognitive impairment in the oldest old," *Neurology*, vol. 50, no. 5, pp. 1496–1498, 1998.
  - [21] J. Verghese, C. Wang, R. B. Lipton, R. Holtzer, and X. Xue, "Quantitative gait dysfunction and risk of cognitive decline and dementia," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 78, no. 9, pp. 929–935, 2007.
  - [22] C. Franceschi, P. Garagnani, P. Parini, C. Giuliani, and A. Santoro, "Inflammaging: a new immune–metabolic viewpoint for age-related diseases," *Nat. Rev. Endocrinol.*, vol. 14, no. 10, pp. 576–590, 2018.
  - [23] S. Leng, P. Chaves, K. Koenig, and J. Walston, "Serum interleukin-6 and hemoglobin as physiological correlates in the geriatric syndrome of frailty: a pilot study," *J. Am. Geriatr. Soc.*, vol. 50, no. 7, pp. 1268–1271, 2002.
  - [24] P. Soysal et al., "Inflammation and frailty in the elderly: A systematic review and meta-analysis," *Aging Res. Rev.*, vol. 31, pp. 1–8, 2016.
  - [25] J. I. Barzilay et al., "Insulin resistance and inflammation as precursors of frailty: the Cardiovascular Health Study: The cardiovascular health study," *Arch. Intern. Med.*, vol. 167, no. 7, pp. 635–641, 2007.
  - [26] L. Fontana et al., "Identification of a metabolic signature for multidimensional impairment and mortality risk in hospitalized older patients," *Aging Cell*, vol. 12, no. 3, pp. 459–466, 2013.
  - [27] C. S. Blaum, Q. L. Xue, J. Tian, R. D. Semba, L. P. Fried, and J. Walston, "Is hyperglycemia associated with frailty status in older women?: Hyperglycemia and frailty in older women," *J. Am. Geriatr. Soc.*, vol. 57, no. 5, pp. 840–847, 2009.
  - [28] R. R. Kalyani et al., "Hyperglycemia and incidence of frailty and lower extremity mobility limitations in older women," *J. Am. Geriatr. Soc.*, vol. 60, no. 9, pp. 1701–1707, 2012.
  - [29] E. García-Esquinas, A. Graciani, P. Guallar-Castillón, E. López-García, L. Rodríguez-Mañas, and F. Rodríguez-Artalejo, "Diabetes and risk of frailty and its

- potential mechanisms: a prospective cohort study of older adults,” *J. Am. Med. Dir. Assoc.*, vol. 16, no. 9, pp. 748–754, 2015.
- [30] D. Khandelwal, A. Goel, U. Kumar, V. Gulati, R. Narang, and A. B. Dey, “Frailty is associated with longer hospital stay and increased mortality in hospitalized older patients,” *J. Nutr. Health Aging*, vol. 16, no. 8, pp. 732–735, 2012.
  - [31] K. Palmer et al., “The relationship between anaemia and frailty: A systematic review and meta-analysis of observational studies,” *J. Nutr. Health Aging*, vol. 22, no. 8, pp. 965–974, 2018.
  - [32] T. B. Vanitallie, “Frailty in the elderly: contributions of sarcopenia and visceral protein depletion,” *Metabolism*, vol. 52, no. 10 Suppl 2, pp. 22–26, 2003.
  - [33] W. R. Hazzard, “Depressed albumin and high-density lipoprotein cholesterol: signposts along the final common pathway of frailty,” *J. Am. Geriatr. Soc.*, vol. 49, no. 9, pp. 1253–1254, 2001.
  - [34] M. S. Chong and S. Sahadevan, “Preclinical Alzheimer’s disease: diagnosis and prediction of progression,” *Lancet Neurol.*, vol. 4, no. 9, pp. 576–579, 2005.
  - [35] S.-K. Herukka, M. Hallikainen, H. Soininen, and T. Pirttilä, “CSF Abeta42 and tau or phosphorylated tau and prediction of progressive mild cognitive impairment,” *Neurology*, vol. 64, no. 7, pp. 1294–1297, 2005.
  - [36] K. Blennow and H. Hampel, “CSF markers for incipient Alzheimer’s disease,” *Lancet Neurol.*, vol. 2, no. 10, pp. 605–613, 2003.
  - [37] V. Van Giau and S. S. A. An, “Optimization of specific multiplex DNA primers to detect variable CLU genomic lesions in patients with Alzheimer’s disease,” *Biochip J.*, vol. 9, no. 4, pp. 278–284, 2015.
  - [38] Z.-X. Yin et al., “High normal plasma triglycerides are associated with preserved cognitive function in Chinese oldest-old,” *Age Aging*, vol. 41, no. 5, pp. 600–606, 2012.
  - [39] Q. He et al., “Relationship between plasma lipids and mild cognitive impairment in the elderly Chinese: a case-control study,” *Lipids Health Dis.*, vol. 15, no. 1, p. 146, 2016.

- [40] S. Stanga, C. Lanni, E. Sinforiani, G. Mazzini, and M. Racchi, “Searching for predictive blood biomarkers: misfolded p53 in mild cognitive impairment,” *Curr. Alzheimer Res.*, vol. 9, no. 10, pp. 1191–1197, 2012.
- [41] F. Babić, L. Trtica Majnarić, S. Bekić, and A. Holzinger, “Machine learning for family doctors: A case of cluster analysis for studying aging associated comorbidities and frailty,” in *Lecture Notes in Computer Science*, Cham: Springer International Publishing, 2019, pp. 178–194,
- [42] M. S. Crane, “Classification and prediction of clinical Alzheimer’s diagnosis based on plasma signalling proteins,” *Ann. Clin. Biochem.*, vol. 45, no. 2, pp. 231–232, 2008.
- [43] A. P. Hassler, E. Menasalvas, F. J. García-García, L. Rodríguez-Mañas, and A. Holzinger, “Importance of medical data preprocessing in predictive modeling and risk factor discovery for the frailty syndrome,” *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, p. 33, 2019.
- [44] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees*. Philadelphia, PA: Chapman & Hall/CRC, 1984.
- [45] B. Krawczyk, “Learning from imbalanced data: open challenges and future directions,” *Prog. Artif. Intell.*, vol. 5, no. 4, pp. 221–232, 2016.
- [46] D.-C. Li, C.-W. Liu, and S. C. Hu, “A learning method for the class imbalance problem with medical data sets,” *Comput. Biol. Med.*, vol. 40, no. 5, pp. 509–518, 2010.
- [47] A. Fernandez, S. Garcia, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, *Learning from imbalanced data sets*. Cham, Switzerland: Springer Nature, 2019.
- [48] W. J. Youden, “Index for rating diagnostic tests,” *Cancer*, vol. 3, no. 1, pp. 32–35, 1950.
- [49] S.-J. Yen and Y.-S. Lee, “Cluster-based under-sampling approaches for imbalanced data distributions,” *Expert Syst. Appl.*, vol. 36, no. 3, pp. 5718–5727, 2009.
- [50] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique,” *arXiv [cs.AI]*, 2011

- [51] C. X. Ling and C. Li, “Data mining for direct marketing: problems and solutions,” 1998, pp. 73–79.
- [52] N. Japkowicz, “Learning from Imbalanced Data Sets: A Comparison of Various Strategies,” 2000.
- [53] L. Ladla and T. Deepa, “Feature Selection Methods And Algorithms,” *International Journal on Computer Science and Engineering (IJCSE)*, vol. 3, no.5, pp. 1787–1797, 2011.
- [54] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene Selection for Cancer Classification using Support Vector Machines. Machine Learning,” *Mach. Learn.*, vol. 46, no. 1/3, pp. 389–422, 2002.
- [55] *MELoR*. [Online]. Available: <https://melor-um.weebly.com/>. [Accessed: 06-Mar-2021].
- [56] M. P. Tan, Y. Y. Ho, A.-V. Chin, N. I. Saedon, I. Z. Abidin, K. H. Chee, H. M. Khor, C. H. Goh, N. N. Hairi, S. Othman, and S. B. Kamaruzzaman, “Ethnic differences in lifetime cumulative incidence of syncope: the Malaysian elders longitudinal research (MELoR) study,” *Clinical Autonomic Research*, vol. 30, no. 2, pp. 121–128, 2019.
- [57] S. Mat, A. H. Razack, J. Lim, S.-Y. Khong, S. B. Kamaruzzaman, A.-V. Chin, A. A. Abbas, N. N. Hairi, S. Othman, and M. P. Tan, “Factors Determining the Increased Risk of Falls in Individuals With Knee Pain in the Malaysian Elders Longitudinal Research (MELoR) Study,” *Frontiers in Medicine*, vol. 6, 2019.
- [58] L. M. Lim, M. Mcstea, W. W. Chung, N. N. Azmi, S. A. A. Aziz, S. Alwi, A. Kamarulzaman, S. B. Kamaruzzaman, S. S. Chua, and R. Rajasuriar, “Prevalence, risk factors and health outcomes associated with polypharmacy among urban community-dwelling older adults in multi-ethnic Malaysia,” *Plos One*, vol. 12, no. 3, 2017.
- [59] J. I. Maletic and A. Marcus, “Data Cleansing,” *Data Mining and Knowledge Discovery Handbook*, pp. 21–36.
- [60] S. Kumar, “7 Ways to Handle Missing Values in Machine Learning,” *Medium*, 02-Aug-2020. [Online]. Available: <https://towardsdatascience.com/7-ways-to-handle-missing-values-in-machine-lea>

- rning-1a6326adf79e. [Accessed: 06-Mar-2021].
- [61] J. Scheffer, “Dealing with Missing Data,” *Research Letters in the Information and Mathematical Sciences*, vol. 3, 2002.
  - [62] I. H. Witten, E. Frank, and M. A. Hall, *Data mining: practical machine learning tools and techniques*. Amsterdam: Elsevier, 2011.
  - [63] K. Muralidharan, “A note on transformation, standardization and normalization,” *International Journal of Operations and Quantitative Management*, vol. 9, 2010.
  - [64] P. J. M. Ali, “Data Normalization and Standardization: A Technical Report,” *Machine Learning Technical Reports*, vol. 1, no. 1, 2014 .
  - [65] S. Lakshmanan, “How, When, and Why Should You Normalize / Standardize / Rescale Your Data?,” *Towards AI - The Best of Tech, Science, and Engineering*, 29-May-2020. [Online]. Available: <https://towardsai.net/p/data-science/how-when-and-why-should-you-normalize-standardize-rescale-your-data-3f083def38ff>. [Accessed: 06-Mar-2021].

## APPENDICES

### APPENDIX A: Criteria Used to Define Frailty

#### *Criteria Used to Define Frailty*

- **Weight loss:** “In the last year, have you lost more than 10 pounds unintentionally (i.e., not due to dieting or exercise)?” If yes, then frail for weight loss criterion. At follow-up, weight loss was calculated as: (Weight in previous year – current measured weight)/(weight in previous year) = K. If  $K \geq 0.05$  and the subject does not report that he/she was trying to lose weight (i.e., unintentional weight loss of at least 5% of previous year’s body weight), then frail for weight loss = Yes.
- **Exhaustion:** Using the CES-D Depression Scale, the following two statements are read. (a) I felt that everything I did was an effort; (b) I could not get going. The question is asked “How often in the last week did you feel this way?” 0 = rarely or none of the time (<1 day), 1 = some or a little of the time (1–2 days), 2 = a moderate amount of the time (3–4 days), or 3 = most of the time. Subjects answering “2” or “3” to either of these questions are categorized as frail by the exhaustion criterion.
- **Physical Activity:** Based on the short version of the Minnesota Leisure Time Activity questionnaire, asking about walking, chores (moderately strenuous), mowing the lawn, raking, gardening, hiking, jogging, biking, exercise cycling, dancing, aerobics, bowling, golf, singles tennis, doubles tennis, racquetball, calisthenics, swimming. Kcals per week expended are calculated using standardized algorithm. This variable is stratified by gender.  
*Men:* Those with Kcals of physical activity per week <383 are frail.  
*Women:* Those with Kcals per week <270 are frail.
- **Walk Time,** stratified by gender and height (gender-specific cutoff a medium height).

<i>Men</i>	<i>Cutoff for Time to Walk 15 feet criterion for frailty</i>
Height $\leq$ 173 cm	$\geq 7$ seconds
Height $>$ 173 cm	$\geq 6$ seconds

<i>Women</i>	<i>Cutoff for Time to Walk 15 feet criterion for frailty</i>
Height $\leq$ 159 cm	$\geq 7$ seconds
Height $>$ 159 cm	$\geq 6$ seconds
- **Grip Strength,** stratified by gender and body mass index (BMI) quartiles:

<i>Men</i>	<i>Cutoff for grip strength (Kg) criterion for frailty</i>
BMI $\leq$ 24	$\leq 29$
BMI 24.1–26	$\leq 30$
BMI 26.1–28	$\leq 30$
BMI $>$ 28	$\leq 32$

<i>Women</i>	<i>Cutoff for grip strength (Kg) criterion for frailty</i>
BMI $\leq$ 23	$\leq 17$
BMI 23.1–26	$\leq 17.3$
BMI 26.1–29	$\leq 18$
BMI $>$ 29	$\leq 21$

## APPENDIX B: Montreal Cognitive Assessment (MoCA)

### MONTREAL COGNITIVE ASSESSMENT (MOCA)

Version 7.1 Original Version

NAME : \_\_\_\_\_  
 Education : \_\_\_\_\_  
 Sex : \_\_\_\_\_  
 Date of birth : \_\_\_\_\_  
 DATE : \_\_\_\_\_

VISUOSPATIAL / EXECUTIVE		Copy cube	Draw CLOCK (Ten past eleven) (3 points)			POINTS		
	[ ]		[ ]	[ ]	[ ]	/5		
			Contour	Numbers	Hands			
NAMING					[ ]	/3		
 MEMORY		FACE	VELVET	CHURCH	DAISY	RED	No points	
1st trial		[ ]	[ ]	[ ]	[ ]	[ ]		
2nd trial		[ ]	[ ]	[ ]	[ ]	[ ]		
ATTENTION		Read list of digits (1 digit/ sec.)	Subject has to repeat them in the forward order [ ] 2 1 8 5 4			/2		
			Subject has to repeat them in the backward order [ ] 7 4 2					
Read list of letters. The subject must tap with his hand at each letter A.		No points if ≥ 2 errors [ ] F B A C M N A A J K L B A F A K D E A A A J A M O F A A B					/1	
Serial 7 subtraction starting at 100		[ ] 93	[ ] 86	[ ] 79	[ ] 72	[ ] 65	/3	
		4 or 5 correct subtractions: 3 pts, 2 or 3 correct: 2 pts, 1 correct: 1 pt, 0 correct: 0 pt						
LANGUAGE		Repeat : I only know that John is the one to help today. [ ] The cat always hid under the couch when dogs were in the room. [ ]					/2	
		Fluency / Name maximum number of words in one minute that begin with the letter F [ ] (N ≥ 11 words)					/1	
ABSTRACTION		Similarity between e.g. banana - orange = fruit [ ] train - bicycle [ ] watch - ruler					/2	
DELAYED RECALL		Has to recall words WITH NO CUE [ ]	FACE [ ]	VELVET [ ]	CHURCH [ ]	DAISY [ ]	RED [ ]	Points for UNCUED recall only /5
Optional		Category cue [ ]	[ ]	[ ]	[ ]	[ ]		
ORIENTATION		[ ] Date	[ ] Month	[ ] Year	[ ] Day	[ ] Place	[ ] City	/6
© Z.Nasreddine MD		www.mocatest.org				Normal ≥ 26 / 30	TOTAL _____/30	
Administered by: _____						Add 1 point if ≤ 12 yr edu		

## APPENDIX C: List of Retained Variables (before Data Pre-processing)

No	Column Name	Blood-based Parameter Name
1	A1_1	Vitamin B12 (pmol/L)
2	A1_2	Serum Folate (nmol/L)
3	A2_1	Serum Homocysteine ( $\mu$ mol/L)
4	A3_1	25-hydroxy Vitamin D (nmol/L)
5	B1_a	Haemoglobin (g/L)
6	B1_a1	RBC (/L)
7	B1_a2	PCV (L/L)
8	B1_a3	MCV (fL)
9	B1_a4	MCH (pg)
10	B1_a5	MCHC (g/L)
11	B1_a6	RDW (%)
12	B1_b"	White Cell Count (/L)
13	B1_b1	Neutrophils (/L)
14	B1_b2	Lymphocytes (/L)
15	B1_b3	Monocytes (/L)
16	B1_b4	Eosinophils (/L)
17	B1_b5	Basophils (/L)
18	B1_c	Platelets (/L)
19	B1_d	Glucose (mmol/L)
20	B2_a1	Total Cholesterol (mmol/L)
21	B2_a2	Triglyceride (mmol/L)
22	B2_a3	HDL Cholesterol (mmol/L)
23	B2_a4	LDL Cholesterol (mmol/L)
24	B2_a5	Total Cholesterol/HDL Ratio
25	B2_b1	Sodium (mmol/L)
26	B2_b2	Potassium (mmol/L)
27	B2_b3	Chloride (mmol/L)
28	B2_c1	Urea (mmol/L)
29	B2_c2	Creatinine (umol/L)
30	B2_c3	eGFR (mL/min/1.73m <sup>2</sup> )
31	B2_c4	Uric Acid (mmol/L)
32	B2_c5	Calcium (mmol/L)
33	B2_c6	Corrected Calcium (mmol/L)
34	B2_c7	Phosphate (mmol/L)
35	B2_d1	Total Protein (g/L)
36	B2_d2	Albumin (g/L)
37	B2_d3	Globulin (g/L)
38	B2_d4	Albumin/Globulin ratio
39	B2_d5	Alkaline Phosphatase (U/L)
40	B2_d6	Total Bilirubin ( $\mu$ mol/L)

41	B2_d7	GGT
42	B2_d8	AST
43	B2_d9	ALT
44	B3	C-Reactive Protein
45	B4_a1	Protein
46	B4_a2	pH
47	B4_a3	Glucose
48	B4_a4	Ketones
49	B4_a5	S.G.
50	B4_a6	Blood
51	B4_b1	Leucocytes (/L)
52	B4_b2	Erythrocytes (/L)
53	B4_b3	Epithelial Cells
54	B5_a1	Free Thyroxine (FT4) (pmol/L)
55	B5_a2	Thyroid Stimulating Hormone (mIU/L)
56	B5_a3	Free Tri-iodothyronine (FT3) (pmol/L)
57	B6	HbA1c

## APPENDIX D: List of Variables after Data Pre-processing

No	Column Name	Blood-based Parameter Name
1	A1_1	Vitamin B12 (pmol/L)
2	A1_2	Serum Folate
3	A2_1	Serum Homocysteine ( $\mu$ mol/L)
4	A3_1	25-hydroxy Vitamin D (nmol/L)
5	B1_a	Haemoglobin (g/L)
6	B1_a1	RBC (/L)
7	B1_a2	PCV (L/L)
8	B1_a3	MCV (fL)
9	B1_a4	MCH (pg)
10	B1_a5	MCHC (g/L)
11	B1_a6	RDW (%)
12	B1_b	White Cell Count (/L)
13	B1_b1	Neutrophils (/L)
14	B1_b2	Lymphocytes (/L)
15	B1_b3	Monocytes (/L)
16	B1_c	Platelets (/L)
17	B1_d	Glucose (mmol/L)
18	B2_a1	Total Cholesterol (mmol/L)
19	B2_a2	Triglyceride (mmol/L)
20	B2_a3	HDL Cholesterol (mmol/L)
21	B2_a4	LDL Cholesterol (mmol/L)
22	B2_a5	Total Cholesterol/HDL Ratio
23	B2_b1	Sodium (mmol/L)
24	B2_b2	Potassium (mmol/L)
25	B2_b3	Chloride (mmol/L)
26	B2_c1	Urea (mmol/L)
27	B2_c2	Creatinine (umol/L)
28	B2_c4	Uric Acid (mmol/L)
29	B2_c5	Calcium (mmol/L)
30	B2_c6	Corrected Calcium (mmol/L)
31	B2_c7	Phosphate (mmol/L)
32	B2_d1	Total Protein (g/L)
33	B2_d2	Albumin (g/L)
34	B2_d3	Globulin (g/L)
35	B2_d4	Albumin/Globulin ratio
36	B2_d5	Alkaline Phosphatase (U/L)
37	B2_d6	Total Bilirubin ( $\mu$ mol/L)
38	B2_d7	GGT
39	B2_d8	AST
40	B2_d9	ALT

41	B3	C-Reactive Protein
42	B4_a2	pH
43	B4_a5	S.G.
44	B5_a2	Thyroid Stimulating Hormone (mIU/L)
45	B5_a3	Free Tri-iodothyronine (FT3) (pmol/L)
46	B6	HbA1c

## APPENDIX E: Confusion Matrices and Performance Evaluation Scores

### 1) Multiclass Classification of 6 Classes – Holdout Method (LOG)

Performance Metrics for Logistic Regression:

0.38177

```
[[104  23  20   0   0   0]
 [ 42  35  12   0   0   0]
 [ 45  19  16   0   0   0]
 [ 26  14  11   0   0   0]
 [ 10  20   4   0   0   0]
 [  3   2   0   0   0   0]]
```

	precision	recall	f1-score	support
0	0.45	0.71	0.55	147
1	0.31	0.39	0.35	89
2	0.25	0.20	0.22	80
3	0.00	0.00	0.00	51
4	0.00	0.00	0.00	34
5	0.00	0.00	0.00	5
accuracy			0.38	406
macro avg	0.17	0.22	0.19	406
weighted avg	0.28	0.38	0.32	406

### 2) Multiclass Classification of 6 Classes – Holdout Method (LDA)

Performance Metrics for LDA:

0.34975

```
[[89  27  26   3   1   1]
 [29  35  17   5   2   1]
 [38  19  17   3   1   2]
 [19  18  11   1   1   1]
 [ 9  17   4   3   0   1]
 [ 3   2   0   0   0   0]]
```

	precision	recall	f1-score	support
0	0.48	0.61	0.53	147
1	0.30	0.39	0.34	89
2	0.23	0.21	0.22	80
3	0.07	0.02	0.03	51
4	0.00	0.00	0.00	34
5	0.00	0.00	0.00	5
accuracy			0.35	406
macro avg	0.18	0.21	0.19	406
weighted avg	0.29	0.35	0.31	406

### 3) Multiclass Classification of 6 Classes – Holdout Method (KNN)

Performance Metrics for KNN:

0.34975

```
[[100  25  16   6   0   0]
 [ 45  29  10   4   1   0]
 [ 45  19  12   4   0   0]
 [ 27  16   6   1   1   0]
 [ 14  14   3   3   0   0]
 [  0   3   1   1   0   0]]
```

	precision	recall	f1-score	support
0	0.43	0.68	0.53	147
1	0.27	0.33	0.30	89
2	0.25	0.15	0.19	80
3	0.05	0.02	0.03	51
4	0.00	0.00	0.00	34
5	0.00	0.00	0.00	5
accuracy			0.35	406
macro avg	0.17	0.20	0.17	406
weighted avg	0.27	0.35	0.30	406

### 4) Multiclass Classification of 6 Classes – Holdout Method (CART)

Performance Metrics for CART:

0.2734

```
[[40  24  49  30   4   0]
 [22  30  19  14   4   0]
 [22  17  27   8   6   0]
 [12   7  17  12   3   0]
 [ 4  16   9   3   2   0]
 [ 1   1   2   0   1   0]]
```

	precision	recall	f1-score	support
0	0.40	0.27	0.32	147
1	0.32	0.34	0.33	89
2	0.22	0.34	0.27	80
3	0.18	0.24	0.20	51
4	0.10	0.06	0.07	34
5	0.00	0.00	0.00	5
accuracy			0.27	406
macro avg	0.20	0.21	0.20	406
weighted avg	0.29	0.27	0.27	406

## 5) Multiclass Classification of 6 Classes – Holdout Method (GNB)

Performance Metrics for GNB:

0.38177

```
[[104 23 20 0 0 0]
 [ 42 35 12 0 0 0]
 [ 45 19 16 0 0 0]
 [ 26 14 11 0 0 0]
 [ 10 20 4 0 0 0]
 [ 3 2 0 0 0 0]]
```

	precision	recall	f1-score	support
0	0.45	0.71	0.55	147
1	0.31	0.39	0.35	89
2	0.25	0.20	0.22	80
3	0.00	0.00	0.00	51
4	0.00	0.00	0.00	34
5	0.00	0.00	0.00	5
accuracy			0.38	406
macro avg	0.17	0.22	0.19	406
weighted avg	0.28	0.38	0.32	406

## 6) Multiclass Classification of 6 Classes – Holdout Method (SVM)

Performance Metrics for SVM:

0.38177

```
[[113 22 12 0 0 0]
 [ 49 33 7 0 0 0]
 [ 54 17 9 0 0 0]
 [ 28 15 8 0 0 0]
 [ 11 20 3 0 0 0]
 [ 3 2 0 0 0 0]]
```

	precision	recall	f1-score	support
0	0.44	0.77	0.56	147
1	0.30	0.37	0.33	89
2	0.23	0.11	0.15	80
3	0.00	0.00	0.00	51
4	0.00	0.00	0.00	34
5	0.00	0.00	0.00	5
accuracy			0.38	406
macro avg	0.16	0.21	0.17	406
weighted avg	0.27	0.38	0.30	406

## 7) Multiclass Classification of 6 Classes – Holdout Method (RFC)

Performance Metrics for RFC:

0.37685

```
[[108 22 17 0 0 0]
 [ 40 32 15 2 0 0]
 [ 47 20 13 0 0 0]
 [ 30 12 9 0 0 0]
 [ 8 19 6 1 0 0]
 [ 2 2 1 0 0 0]]
```

	precision	recall	f1-score	support
0	0.46	0.73	0.57	147
1	0.30	0.36	0.33	89
2	0.21	0.16	0.18	80
3	0.00	0.00	0.00	51
4	0.00	0.00	0.00	34
5	0.00	0.00	0.00	5
accuracy			0.38	406
macro avg	0.16	0.21	0.18	406
weighted avg	0.27	0.38	0.31	406

## 8) Multiclass Classification of 6 Classes – 5-fold Cross Validation (LOG)

	precision	recall	f1-score	support
0	0.44	0.91	0.59	68
1	0.52	0.52	0.52	46
2	0.40	0.13	0.20	45
3	0.00	0.00	0.00	27
4	0.00	0.00	0.00	15
5	0.00	0.00	0.00	2
accuracy			0.45	203
macro avg	0.23	0.26	0.22	203
weighted avg	0.35	0.45	0.36	203
	precision	recall	f1-score	support
0	0.41	0.69	0.51	68
1	0.34	0.43	0.38	47
2	0.16	0.09	0.11	45
3	0.50	0.04	0.07	26
4	0.00	0.00	0.00	15
5	0.00	0.00	0.00	2
accuracy			0.35	203
macro avg	0.24	0.21	0.18	203
weighted avg	0.32	0.35	0.29	203
	precision	recall	f1-score	support
0	0.48	0.75	0.58	69
1	0.33	0.47	0.39	47
2	0.24	0.11	0.15	44
3	0.17	0.04	0.06	26
4	0.00	0.00	0.00	16
5	0.00	0.00	0.00	1
accuracy			0.39	203
macro avg	0.20	0.23	0.20	203
weighted avg	0.31	0.39	0.33	203

	precision	recall	f1-score	support
0	0.45	0.87	0.59	69
1	0.38	0.45	0.41	47
2	0.18	0.05	0.07	44
3	1.00	0.04	0.07	27
4	0.00	0.00	0.00	15
5	0.00	0.00	0.00	1
accuracy			0.41	203
macro avg	0.34	0.23	0.19	203
weighted avg	0.41	0.41	0.32	203

	precision	recall	f1-score	support
0	0.36	0.70	0.48	69
1	0.25	0.28	0.27	46
2	0.25	0.09	0.13	45
3	0.33	0.04	0.07	27
4	0.00	0.00	0.00	15
5	0.00	0.00	0.00	1
accuracy			0.33	203
macro avg	0.20	0.18	0.16	203
weighted avg	0.28	0.33	0.26	203

[0.45320197 0.3546798 0.39408867 0.4137931 0.32512315]

### 9) Multiclass Classification of 6 Classes – 5-fold Cross Validation (LDA)

	precision	recall	f1-score	support
0	0.44	0.79	0.57	68
1	0.44	0.50	0.47	46
2	0.29	0.16	0.20	45
3	0.00	0.00	0.00	27
4	0.00	0.00	0.00	15
5	0.00	0.00	0.00	2
accuracy			0.41	203
macro avg	0.20	0.24	0.21	203
weighted avg	0.31	0.41	0.34	203

	precision	recall	f1-score	support
0	0.45	0.63	0.52	68
1	0.39	0.47	0.43	47
2	0.14	0.09	0.11	45
3	0.12	0.04	0.06	26
4	0.00	0.00	0.00	15
5	0.00	0.00	0.00	2
accuracy			0.34	203
macro avg	0.18	0.20	0.19	203
weighted avg	0.29	0.34	0.31	203

	precision	recall	f1-score	support
0	0.48	0.64	0.55	69
1	0.25	0.34	0.29	47
2	0.32	0.18	0.23	44
3	0.25	0.12	0.16	26
4	0.00	0.00	0.00	16
5	0.00	0.00	0.00	1
accuracy			0.35	203
macro avg	0.22	0.21	0.20	203
weighted avg	0.32	0.35	0.32	203

	precision	recall	f1-score	support
0	0.43	0.72	0.54	69
1	0.42	0.51	0.46	47
2	0.15	0.07	0.09	44
3	0.25	0.04	0.06	27
4	0.00	0.00	0.00	15
5	0.00	0.00	0.00	1
accuracy			0.38	203
macro avg	0.21	0.22	0.19	203
weighted avg	0.31	0.38	0.32	203

	precision	recall	f1-score	support
0	0.36	0.62	0.46	69
1	0.26	0.26	0.26	46
2	0.32	0.13	0.19	45
3	0.27	0.11	0.16	27
4	0.12	0.07	0.09	15
5	0.00	0.00	0.00	1
accuracy			0.32	203
macro avg	0.22	0.20	0.19	203
weighted avg	0.30	0.32	0.28	203

## 10) Multiclass Classification of 6 Classes – 5-fold Cross Validation (KNN)

	precision	recall	f1-score	support
0	0.41	0.76	0.53	68
1	0.33	0.35	0.34	46
2	0.12	0.04	0.07	45
3	0.00	0.00	0.00	27
4	0.00	0.00	0.00	15
5	0.00	0.00	0.00	2
accuracy			0.34	203
macro avg	0.14	0.19	0.16	203
weighted avg	0.24	0.34	0.27	203
	precision	recall	f1-score	support
0	0.34	0.59	0.43	68
1	0.21	0.19	0.20	47
2	0.19	0.13	0.16	45
3	0.00	0.00	0.00	26
4	0.00	0.00	0.00	15
5	0.00	0.00	0.00	2
accuracy			0.27	203
macro avg	0.12	0.15	0.13	203
weighted avg	0.21	0.27	0.23	203
	precision	recall	f1-score	support
0	0.39	0.54	0.45	69
1	0.37	0.43	0.40	47
2	0.21	0.18	0.19	44
3	0.23	0.12	0.15	26
4	1.00	0.06	0.12	16
5	0.00	0.00	0.00	1
accuracy			0.34	203
macro avg	0.37	0.22	0.22	203
weighted avg	0.37	0.34	0.31	203

	precision	recall	f1-score	support
0	0.36	0.65	0.46	69
1	0.37	0.32	0.34	47
2	0.12	0.07	0.09	44
3	0.00	0.00	0.00	27
4	0.00	0.00	0.00	15
5	0.00	0.00	0.00	1
accuracy			0.31	203
macro avg	0.14	0.17	0.15	203
weighted avg	0.23	0.31	0.26	203

	precision	recall	f1-score	support
0	0.35	0.59	0.44	69
1	0.30	0.33	0.31	46
2	0.18	0.09	0.12	45
3	0.15	0.07	0.10	27
4	0.00	0.00	0.00	15
5	0.00	0.00	0.00	1
accuracy			0.31	203
macro avg	0.16	0.18	0.16	203
weighted avg	0.25	0.31	0.26	203

## 11) Multiclass Classification of 6 Classes – 5-fold Cross Validation (CART)

	precision	recall	f1-score	support
0	0.37	0.44	0.40	68
1	0.27	0.26	0.26	46
2	0.21	0.18	0.19	45
3	0.14	0.11	0.12	27
4	0.06	0.07	0.06	15
5	0.00	0.00	0.00	2
accuracy			0.27	203
macro avg	0.17	0.18	0.17	203
weighted avg	0.25	0.27	0.26	203

	precision	recall	f1-score	support
0	0.41	0.35	0.38	68
1	0.19	0.19	0.19	47
2	0.24	0.24	0.24	45
3	0.04	0.04	0.04	26
4	0.08	0.13	0.10	15
5	0.00	0.00	0.00	2
accuracy			0.23	203
macro avg	0.16	0.16	0.16	203
weighted avg	0.24	0.23	0.24	203

	precision	recall	f1-score	support
0	0.41	0.41	0.41	69
1	0.28	0.28	0.28	47
2	0.28	0.27	0.28	44
3	0.11	0.08	0.09	26
4	0.10	0.12	0.11	16
5	0.00	0.00	0.00	1
accuracy			0.28	203
macro avg	0.19	0.19	0.19	203
weighted avg	0.28	0.28	0.28	203
	precision	recall	f1-score	support
0	0.41	0.41	0.41	69
1	0.25	0.30	0.27	47
2	0.23	0.18	0.20	44
3	0.24	0.19	0.21	27
4	0.09	0.13	0.11	15
5	0.00	0.00	0.00	1
accuracy			0.28	203
macro avg	0.20	0.20	0.20	203
weighted avg	0.29	0.28	0.28	203
	precision	recall	f1-score	support
0	0.38	0.46	0.42	69
1	0.22	0.15	0.18	46
2	0.26	0.22	0.24	45
3	0.22	0.22	0.22	27
4	0.26	0.33	0.29	15
5	0.00	0.00	0.00	1
accuracy			0.30	203
macro avg	0.22	0.23	0.23	203
weighted avg	0.28	0.30	0.29	203

## 12) Multiclass Classification of 6 Classes – 5-fold Cross Validation (GNB)

	precision	recall	f1-score	support
0	0.43	0.66	0.52	68
1	0.38	0.24	0.29	46
2	0.32	0.16	0.21	45
3	0.22	0.30	0.25	27
4	0.22	0.13	0.17	15
5	0.00	0.00	0.00	2
accuracy			0.36	203
macro avg	0.26	0.25	0.24	203
weighted avg	0.35	0.36	0.33	203

	precision	recall	f1-score	support
0	0.38	0.53	0.44	68
1	0.38	0.26	0.30	47
2	0.11	0.04	0.06	45
3	0.20	0.27	0.23	26
4	0.15	0.20	0.17	15
5	0.00	0.00	0.00	2
accuracy			0.30	203
macro avg	0.20	0.22	0.20	203
weighted avg	0.27	0.30	0.27	203
	precision	recall	f1-score	support
0	0.42	0.62	0.50	69
1	0.19	0.17	0.18	47
2	0.12	0.02	0.04	44
3	0.19	0.27	0.23	26
4	0.15	0.12	0.14	15
5	0.00	0.00	0.00	1
accuracy			0.30	203
macro avg	0.18	0.20	0.18	203
weighted avg	0.25	0.30	0.26	203
	precision	recall	f1-score	support
0	0.44	0.71	0.54	69
1	0.35	0.30	0.32	47
2	0.46	0.14	0.21	44
3	0.15	0.15	0.15	27
4	0.20	0.13	0.16	15
5	0.00	0.00	0.00	1
accuracy			0.37	203
macro avg	0.27	0.24	0.23	203
weighted avg	0.37	0.37	0.34	203
	precision	recall	f1-score	support
0	0.42	0.75	0.54	69
1	0.24	0.11	0.15	46
2	0.33	0.11	0.17	45
3	0.09	0.07	0.08	27
4	0.21	0.20	0.21	15
5	0.00	0.00	0.00	1
accuracy			0.33	203
macro avg	0.22	0.21	0.19	203
weighted avg	0.30	0.33	0.28	203

### 13) Multiclass Classification of 6 Classes – 5-fold Cross Validation (SVM)

	precision	recall	f1-score	support
0	0.41	0.93	0.57	68
1	0.49	0.50	0.49	46
2	0.00	0.00	0.00	45
3	0.00	0.00	0.00	27
4	0.00	0.00	0.00	15
5	0.00	0.00	0.00	2
accuracy			0.42	203
macro avg	0.15	0.24	0.18	203
weighted avg	0.25	0.42	0.30	203

	precision	recall	f1-score	support
0	0.41	0.76	0.53	68
1	0.33	0.43	0.37	47
2	0.07	0.02	0.03	45
3	0.00	0.00	0.00	26
4	0.00	0.00	0.00	15
5	0.00	0.00	0.00	2
accuracy			0.36	203
macro avg	0.13	0.20	0.16	203
weighted avg	0.23	0.36	0.27	203

	precision	recall	f1-score	support
0	0.45	0.80	0.58	69
1	0.36	0.55	0.43	47
2	0.25	0.05	0.08	44
3	0.00	0.00	0.00	26
4	0.00	0.00	0.00	16
5	0.00	0.00	0.00	1
accuracy			0.41	203
macro avg	0.18	0.23	0.18	203
weighted avg	0.29	0.41	0.31	203

	precision	recall	f1-score	support
0	0.42	0.91	0.58	69
1	0.42	0.47	0.44	47
2	0.50	0.02	0.04	44
3	0.00	0.00	0.00	27
4	0.00	0.00	0.00	15
5	0.00	0.00	0.00	1
accuracy			0.42	203
macro avg	0.22	0.23	0.18	203
weighted avg	0.35	0.42	0.31	203

	precision	recall	f1-score	support
0	0.37	0.78	0.50	69
1	0.26	0.30	0.28	46
2	0.00	0.00	0.00	45
3	0.00	0.00	0.00	27
4	0.00	0.00	0.00	15
5	0.00	0.00	0.00	1
accuracy			0.33	203
macro avg	0.11	0.18	0.13	203
weighted avg	0.19	0.33	0.23	203

#### 14) Multiclass Classification of 6 Classes – 5-fold Cross Validation (RFC)

	precision	recall	f1-score	support
0	0.39	0.82	0.53	68
1	0.40	0.30	0.35	46
2	0.23	0.11	0.15	45
3	0.33	0.04	0.07	27
4	0.00	0.00	0.00	15
5	0.00	0.00	0.00	2
accuracy			0.37	203
macro avg	0.23	0.21	0.18	203
weighted avg	0.32	0.37	0.30	203

	precision	recall	f1-score	support
0	0.40	0.68	0.50	68
1	0.33	0.45	0.38	47
2	0.10	0.04	0.06	45
3	0.00	0.00	0.00	26
4	0.50	0.07	0.12	15
5	0.00	0.00	0.00	2
accuracy			0.34	203
macro avg	0.22	0.21	0.18	203
weighted avg	0.27	0.34	0.28	203

	precision	recall	f1-score	support
0	0.41	0.71	0.52	69
1	0.29	0.40	0.34	47
2	0.24	0.09	0.13	44
3	0.00	0.00	0.00	26
4	0.00	0.00	0.00	16
5	0.00	0.00	0.00	1
accuracy			0.35	203
macro avg	0.16	0.20	0.16	203
weighted avg	0.26	0.35	0.28	203

	precision	recall	f1-score	support
0	0.43	0.81	0.57	69
1	0.36	0.40	0.38	47
2	0.11	0.05	0.06	44
3	1.00	0.04	0.07	27
4	0.00	0.00	0.00	15
5	0.00	0.00	0.00	1
accuracy			0.38	203
macro avg	0.32	0.22	0.18	203
weighted avg	0.39	0.38	0.30	203
	precision	recall	f1-score	support
0	0.40	0.83	0.54	69
1	0.32	0.28	0.30	46
2	0.38	0.13	0.20	45
3	0.00	0.00	0.00	27
4	1.00	0.07	0.12	15
5	0.00	0.00	0.00	1
accuracy			0.38	203
macro avg	0.35	0.22	0.19	203
weighted avg	0.37	0.38	0.30	203

## 15) Binary Classification of Robust and Non-Robust – Holdout Method (LOG)

Performance Metrics for Logistic Regression:

0.66502

[[249 10]  
[126 21]]

	precision	recall	f1-score	support
0	0.66	0.96	0.79	259
1	0.68	0.14	0.24	147
accuracy			0.67	406
macro avg	0.67	0.55	0.51	406
weighted avg	0.67	0.67	0.59	406

## 16) Binary Classification of Robust and Non-Robust – Holdout Method (LDA)

Performance Metrics for LDA:

0.65517

[[224 35]  
[105 42]]

	precision	recall	f1-score	support
0	0.68	0.86	0.76	259
1	0.55	0.29	0.37	147
accuracy			0.66	406
macro avg	0.61	0.58	0.57	406
weighted avg	0.63	0.66	0.62	406

## 17) Binary Classification of Robust and Non-Robust – Holdout Method (KNN)

Performance Metrics for KNN:

0.59852

[[192 67]  
[ 96 51]]

	precision	recall	f1-score	support
0	0.67	0.74	0.70	259
1	0.43	0.35	0.38	147
accuracy			0.60	406
macro avg	0.55	0.54	0.54	406
weighted avg	0.58	0.60	0.59	406

## 18) Binary Classification of Robust and Non-Robust – Holdout Method (CART)

Performance Metrics for CART:

0.55911

[[171 88]  
[ 91 56]]

	precision	recall	f1-score	support
0	0.65	0.66	0.66	259
1	0.39	0.38	0.38	147
accuracy			0.56	406
macro avg	0.52	0.52	0.52	406
weighted avg	0.56	0.56	0.56	406

### 19) Binary Classification of Robust and Non-Robust – Holdout Method (GNB)

Performance Metrics for GNB:

0.66502

[[249 10]  
[126 21]]

	precision	recall	f1-score	support
0	0.66	0.96	0.79	259
1	0.68	0.14	0.24	147
accuracy			0.67	406
macro avg	0.67	0.55	0.51	406
weighted avg	0.67	0.67	0.59	406

### 20) Binary Classification of Robust and Non-Robust – Holdout Method (SVM)

Performance Metrics for SVM:

0.63793

[[259 0]  
[147 0]]

	precision	recall	f1-score	support
0	0.64	1.00	0.78	259
1	0.00	0.00	0.00	147
accuracy			0.64	406
macro avg	0.32	0.50	0.39	406
weighted avg	0.41	0.64	0.50	406

### 21) Binary Classification of Robust and Non-Robust – Holdout Method (RFC)

Performance Metrics for RFC:

0.64532

[[246 13]  
[131 16]]

	precision	recall	f1-score	support
0	0.65	0.95	0.77	259
1	0.55	0.11	0.18	147
accuracy			0.65	406
macro avg	0.60	0.53	0.48	406
weighted avg	0.62	0.65	0.56	406

## 22) Binary Classification of Robust and Non-Robust – 5-fold Cross Validation (LOG)

	precision	recall	f1-score	support
0	0.68	0.98	0.80	135
1	0.70	0.10	0.18	68
accuracy			0.68	203
macro avg	0.69	0.54	0.49	203
weighted avg	0.69	0.68	0.60	203
	precision	recall	f1-score	support
0	0.71	0.96	0.81	135
1	0.71	0.22	0.34	68
accuracy			0.71	203
macro avg	0.71	0.59	0.58	203
weighted avg	0.71	0.71	0.65	203
	precision	recall	f1-score	support
0	0.67	0.98	0.80	134
1	0.62	0.07	0.13	69
accuracy			0.67	203
macro avg	0.65	0.53	0.46	203
weighted avg	0.66	0.67	0.57	203
	precision	recall	f1-score	support
0	0.68	0.87	0.76	134
1	0.43	0.19	0.26	69
accuracy			0.64	203
macro avg	0.55	0.53	0.51	203
weighted avg	0.59	0.64	0.59	203
	precision	recall	f1-score	support
0	0.65	0.81	0.72	134
1	0.28	0.14	0.19	69
accuracy			0.58	203
macro avg	0.46	0.48	0.45	203
weighted avg	0.52	0.58	0.54	203

### 23) Binary Classification of Robust and Non-Robust – 5-fold Cross Validation (LDA)

	precision	recall	f1-score	support
0	0.70	0.87	0.78	135
1	0.51	0.26	0.35	68
accuracy			0.67	203
macro avg	0.61	0.57	0.56	203
weighted avg	0.64	0.67	0.64	203
	precision	recall	f1-score	support
0	0.72	0.90	0.80	135
1	0.61	0.32	0.42	68
accuracy			0.70	203
macro avg	0.67	0.61	0.61	203
weighted avg	0.69	0.70	0.67	203
	precision	recall	f1-score	support
0	0.69	0.96	0.80	134
1	0.67	0.17	0.28	69
accuracy			0.69	203
macro avg	0.68	0.56	0.54	203
weighted avg	0.68	0.69	0.62	203
	precision	recall	f1-score	support
0	0.71	0.81	0.75	134
1	0.48	0.35	0.40	69
accuracy			0.65	203
macro avg	0.59	0.58	0.58	203
weighted avg	0.63	0.65	0.63	203
	precision	recall	f1-score	support
0	0.66	0.72	0.69	134
1	0.33	0.28	0.30	69
accuracy			0.57	203
macro avg	0.50	0.50	0.49	203
weighted avg	0.55	0.57	0.56	203

## 24) Binary Classification of Robust and Non-Robust – 5-fold Cross Validation (KNN)

	precision	recall	f1-score	support
0	0.73	0.79	0.76	135
1	0.51	0.43	0.46	68
accuracy			0.67	203
macro avg	0.62	0.61	0.61	203
weighted avg	0.66	0.67	0.66	203
	precision	recall	f1-score	support
0	0.70	0.78	0.74	135
1	0.43	0.34	0.38	68
accuracy			0.63	203
macro avg	0.57	0.56	0.56	203
weighted avg	0.61	0.63	0.62	203
	precision	recall	f1-score	support
0	0.66	0.75	0.70	134
1	0.33	0.23	0.27	69
accuracy			0.58	203
macro avg	0.49	0.49	0.49	203
weighted avg	0.54	0.58	0.56	203
	precision	recall	f1-score	support
0	0.70	0.67	0.68	134
1	0.41	0.43	0.42	69
accuracy			0.59	203
macro avg	0.55	0.55	0.55	203
weighted avg	0.60	0.59	0.59	203
	precision	recall	f1-score	support
0	0.65	0.66	0.66	134
1	0.33	0.32	0.32	69
accuracy			0.55	203
macro avg	0.49	0.49	0.49	203
weighted avg	0.54	0.55	0.55	203

## 25) Binary Classification of Robust and Non-Robust – 5-fold Cross Validation (CART)

	precision	recall	f1-score	support
0	0.68	0.68	0.68	135
1	0.36	0.35	0.36	68
accuracy			0.57	203
macro avg	0.52	0.52	0.52	203
weighted avg	0.57	0.57	0.57	203
	precision	recall	f1-score	support
0	0.70	0.78	0.74	135
1	0.44	0.35	0.39	68
accuracy			0.64	203
macro avg	0.57	0.57	0.57	203
weighted avg	0.62	0.64	0.62	203
	precision	recall	f1-score	support
0	0.73	0.74	0.74	134
1	0.49	0.48	0.48	69
accuracy			0.65	203
macro avg	0.61	0.61	0.61	203
weighted avg	0.65	0.65	0.65	203
	precision	recall	f1-score	support
0	0.61	0.57	0.59	134
1	0.26	0.29	0.27	69
accuracy			0.48	203
macro avg	0.44	0.43	0.43	203
weighted avg	0.49	0.48	0.48	203
	precision	recall	f1-score	support
0	0.69	0.66	0.68	134
1	0.39	0.42	0.41	69
accuracy			0.58	203
macro avg	0.54	0.54	0.54	203
weighted avg	0.59	0.58	0.58	203

## 26) Binary Classification of Robust and Non-Robust – 5-fold Cross Validation (GNB)

	precision	recall	f1-score	support
0	0.86	0.44	0.59	135
1	0.44	0.85	0.58	68
accuracy			0.58	203
macro avg	0.65	0.65	0.58	203
weighted avg	0.72	0.58	0.58	203
	precision	recall	f1-score	support
0	0.75	0.27	0.39	135
1	0.36	0.82	0.50	68
accuracy			0.45	203
macro avg	0.56	0.55	0.45	203
weighted avg	0.62	0.45	0.43	203
	precision	recall	f1-score	support
0	0.73	0.34	0.46	134
1	0.37	0.75	0.50	69
accuracy			0.48	203
macro avg	0.55	0.54	0.48	203
weighted avg	0.60	0.48	0.47	203
	precision	recall	f1-score	support
0	0.78	0.43	0.55	134
1	0.41	0.77	0.53	69
accuracy			0.54	203
macro avg	0.59	0.60	0.54	203
weighted avg	0.65	0.54	0.54	203
	precision	recall	f1-score	support
0	0.73	0.18	0.29	134
1	0.35	0.87	0.50	69
accuracy			0.41	203
macro avg	0.54	0.52	0.39	203
weighted avg	0.60	0.41	0.36	203

## 27) Binary Classification of Robust and Non-Robust – 5-fold Cross Validation (SVM)

	precision	recall	f1-score	support
0	0.67	1.00	0.80	135
1	0.00	0.00	0.00	68
accuracy			0.67	203
macro avg	0.33	0.50	0.40	203
weighted avg	0.44	0.67	0.53	203
	precision	recall	f1-score	support
0	0.67	1.00	0.80	135
1	0.00	0.00	0.00	68
accuracy			0.67	203
macro avg	0.33	0.50	0.40	203
weighted avg	0.44	0.67	0.53	203
	precision	recall	f1-score	support
0	0.66	1.00	0.80	134
1	0.00	0.00	0.00	69
accuracy			0.66	203
macro avg	0.33	0.50	0.40	203
weighted avg	0.44	0.66	0.52	203
	precision	recall	f1-score	support
0	0.66	1.00	0.80	134
1	0.00	0.00	0.00	69
accuracy			0.66	203
macro avg	0.33	0.50	0.40	203
weighted avg	0.44	0.66	0.52	203
	precision	recall	f1-score	support
0	0.64	0.87	0.74	134
1	0.22	0.07	0.11	69
accuracy			0.60	203
macro avg	0.43	0.47	0.42	203
weighted avg	0.50	0.60	0.52	203

## 28) Binary Classification of Robust and Non-Robust – 5-fold Cross Validation (RFC)

	precision	recall	f1-score	support
0	0.68	0.98	0.80	135
1	0.62	0.07	0.13	68
accuracy			0.67	203
macro avg	0.65	0.53	0.47	203
weighted avg	0.66	0.67	0.58	203
	precision	recall	f1-score	support
0	0.67	0.94	0.78	135
1	0.38	0.07	0.12	68
accuracy			0.65	203
macro avg	0.53	0.51	0.45	203
weighted avg	0.57	0.65	0.56	203
	precision	recall	f1-score	support
0	0.68	0.96	0.80	134
1	0.62	0.12	0.20	69
accuracy			0.67	203
macro avg	0.65	0.54	0.50	203
weighted avg	0.66	0.67	0.59	203
	precision	recall	f1-score	support
0	0.69	0.91	0.78	134
1	0.52	0.19	0.28	69
accuracy			0.67	203
macro avg	0.60	0.55	0.53	203
weighted avg	0.63	0.67	0.61	203
	precision	recall	f1-score	support
0	0.69	0.90	0.78	134
1	0.52	0.20	0.29	69
accuracy			0.67	203
macro avg	0.60	0.55	0.54	203
weighted avg	0.63	0.67	0.61	203

29) Binary Classification of Robust and Frail with MCI (76 samples) – Holdout Method  
(LOG)

Performance Metrics for Logistic Regression:

0.55738

[[18 12]  
[15 16]]

	precision	recall	f1-score	support
0	0.55	0.60	0.57	30
1	0.57	0.52	0.54	31
accuracy			0.56	61
macro avg	0.56	0.56	0.56	61
weighted avg	0.56	0.56	0.56	61

30) Binary Classification of Robust and Frail with MCI (76 samples) – Holdout Method  
(LDA)

Performance Metrics for LDA:

0.67213

[[21 9]  
[11 20]]

	precision	recall	f1-score	support
0	0.66	0.70	0.68	30
1	0.69	0.65	0.67	31
accuracy			0.67	61
macro avg	0.67	0.67	0.67	61
weighted avg	0.67	0.67	0.67	61

31) Binary Classification of Robust and Frail with MCI (76 samples) – Holdout Method  
(KNN)

```
Performance Metrics for KNN:  
0.54098  
[[25 5]  
 [23 8]]  


|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.52      | 0.83   | 0.64     | 30      |
| 1            | 0.62      | 0.26   | 0.36     | 31      |
| accuracy     |           |        | 0.54     | 61      |
| macro avg    | 0.57      | 0.55   | 0.50     | 61      |
| weighted avg | 0.57      | 0.54   | 0.50     | 61      |


```

32) Binary Classification of Robust and Frail with MCI (76 samples) – Holdout Method  
(CART)

```
Performance Metrics for CART:  
0.62295  
[[18 12]  
 [11 20]]  


|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.62      | 0.60   | 0.61     | 30      |
| 1            | 0.62      | 0.65   | 0.63     | 31      |
| accuracy     |           |        | 0.62     | 61      |
| macro avg    | 0.62      | 0.62   | 0.62     | 61      |
| weighted avg | 0.62      | 0.62   | 0.62     | 61      |


```

33) Binary Classification of Robust and Frail with MCI (76 samples) – Holdout Method  
(GNB)

Performance Metrics for GNB:

0.55738

[[18 12]  
[15 16]]

	precision	recall	f1-score	support
0	0.55	0.60	0.57	30
1	0.57	0.52	0.54	31
accuracy			0.56	61
macro avg	0.56	0.56	0.56	61
weighted avg	0.56	0.56	0.56	61

34) Binary Classification of Robust and Frail with MCI (76 samples) – Holdout Method  
(SVM)

Performance Metrics for SVM:

0.63934

[[20 10]  
[12 19]]

	precision	recall	f1-score	support
0	0.62	0.67	0.65	30
1	0.66	0.61	0.63	31
accuracy			0.64	61
macro avg	0.64	0.64	0.64	61
weighted avg	0.64	0.64	0.64	61

35) Binary Classification of Robust and Frail with MCI (76 samples) – Holdout Method (RFC)

```
Performance Metrics for RFC:  
0.62295  
[[20 10]  
 [13 18]]  


|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.61      | 0.67   | 0.63     | 30      |
| 1            | 0.64      | 0.58   | 0.61     | 31      |
| accuracy     |           |        | 0.62     | 61      |
| macro avg    | 0.62      | 0.62   | 0.62     | 61      |
| weighted avg | 0.62      | 0.62   | 0.62     | 61      |


```

36) Binary Classification of Robust and Frail with MCI (76 samples) – 5-fold Cross Validation (LOG)

```


|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.92      | 0.69   | 0.79     | 16      |
| 1            | 0.74      | 0.93   | 0.82     | 15      |
| accuracy     |           |        | 0.81     | 31      |
| macro avg    | 0.83      | 0.81   | 0.80     | 31      |
| weighted avg | 0.83      | 0.81   | 0.80     | 31      |


|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.62      | 0.67   | 0.65     | 15      |
| 1            | 0.67      | 0.62   | 0.65     | 16      |
| accuracy     |           |        | 0.65     | 31      |
| macro avg    | 0.65      | 0.65   | 0.65     | 31      |
| weighted avg | 0.65      | 0.65   | 0.65     | 31      |


|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.80      | 0.80   | 0.80     | 15      |
| 1            | 0.80      | 0.80   | 0.80     | 15      |
| accuracy     |           |        | 0.80     | 30      |
| macro avg    | 0.80      | 0.80   | 0.80     | 30      |
| weighted avg | 0.80      | 0.80   | 0.80     | 30      |


```

	precision	recall	f1-score	support
0	0.73	0.73	0.73	15
1	0.73	0.73	0.73	15
accuracy			0.73	30
macro avg	0.73	0.73	0.73	30
weighted avg	0.73	0.73	0.73	30
	precision	recall	f1-score	support
0	0.65	0.73	0.69	15
1	0.69	0.60	0.64	15
accuracy			0.67	30
macro avg	0.67	0.67	0.67	30
weighted avg	0.67	0.67	0.67	30

37) Binary Classification of Robust and Frail with MCI (76 samples) – 5-fold Cross Validation (LDA)

	precision	recall	f1-score	support
0	0.69	0.56	0.62	16
1	0.61	0.73	0.67	15
accuracy			0.65	31
macro avg	0.65	0.65	0.64	31
weighted avg	0.65	0.65	0.64	31
	precision	recall	f1-score	support
0	0.60	0.60	0.60	15
1	0.62	0.62	0.62	16
accuracy			0.61	31
macro avg	0.61	0.61	0.61	31
weighted avg	0.61	0.61	0.61	31
	precision	recall	f1-score	support
0	0.67	0.80	0.73	15
1	0.75	0.60	0.67	15
accuracy			0.70	30
macro avg	0.71	0.70	0.70	30
weighted avg	0.71	0.70	0.70	30

	precision	recall	f1-score	support
0	0.59	0.67	0.62	15
1	0.62	0.53	0.57	15
accuracy			0.60	30
macro avg	0.60	0.60	0.60	30
weighted avg	0.60	0.60	0.60	30
	precision	recall	f1-score	support
0	0.67	0.67	0.67	15
1	0.67	0.67	0.67	15
accuracy			0.67	30
macro avg	0.67	0.67	0.67	30
weighted avg	0.67	0.67	0.67	30

38) Binary Classification of Robust and Frail with MCI (76 samples) – 5-fold Cross Validation (KNN)

	precision	recall	f1-score	support
0	0.70	0.88	0.78	16
1	0.82	0.60	0.69	15
accuracy			0.74	31
macro avg	0.76	0.74	0.74	31
weighted avg	0.76	0.74	0.74	31
	precision	recall	f1-score	support
0	0.62	0.87	0.72	15
1	0.80	0.50	0.62	16
accuracy			0.68	31
macro avg	0.71	0.68	0.67	31
weighted avg	0.71	0.68	0.67	31
	precision	recall	f1-score	support
0	0.60	0.80	0.69	15
1	0.70	0.47	0.56	15
accuracy			0.63	30
macro avg	0.65	0.63	0.62	30
weighted avg	0.65	0.63	0.62	30

	precision	recall	f1-score	support
0	0.57	0.80	0.67	15
1	0.67	0.40	0.50	15
accuracy			0.60	30
macro avg	0.62	0.60	0.58	30
weighted avg	0.62	0.60	0.58	30
	precision	recall	f1-score	support
0	0.64	0.93	0.76	15
1	0.88	0.47	0.61	15
accuracy			0.70	30
macro avg	0.76	0.70	0.68	30
weighted avg	0.76	0.70	0.68	30

39) Binary Classification of Robust and Frail with MCI (76 samples) – 5-fold Cross

Validation (CART)

	precision	recall	f1-score	support
0	0.47	0.44	0.45	16
1	0.44	0.47	0.45	15
accuracy			0.45	31
macro avg	0.45	0.45	0.45	31
weighted avg	0.45	0.45	0.45	31
	precision	recall	f1-score	support
0	0.60	0.40	0.48	15
1	0.57	0.75	0.65	16
accuracy			0.58	31
macro avg	0.59	0.57	0.56	31
weighted avg	0.59	0.58	0.57	31
	precision	recall	f1-score	support
0	0.69	0.60	0.64	15
1	0.65	0.73	0.69	15
accuracy			0.67	30
macro avg	0.67	0.67	0.67	30
weighted avg	0.67	0.67	0.67	30

	precision	recall	f1-score	support
0	0.47	0.47	0.47	15
1	0.47	0.47	0.47	15
accuracy			0.47	30
macro avg	0.47	0.47	0.47	30
weighted avg	0.47	0.47	0.47	30
	precision	recall	f1-score	support
0	0.53	0.53	0.53	15
1	0.53	0.53	0.53	15
accuracy			0.53	30
macro avg	0.53	0.53	0.53	30
weighted avg	0.53	0.53	0.53	30

40) Binary Classification of Robust and Frail with MCI (76 samples) – 5-fold Cross Validation (GNB)

	precision	recall	f1-score	support
0	0.72	0.81	0.76	16
1	0.77	0.67	0.71	15
accuracy			0.74	31
macro avg	0.75	0.74	0.74	31
weighted avg	0.74	0.74	0.74	31
	precision	recall	f1-score	support
0	0.70	0.93	0.80	15
1	0.91	0.62	0.74	16
accuracy			0.77	31
macro avg	0.80	0.78	0.77	31
weighted avg	0.81	0.77	0.77	31

	precision	recall	f1-score	support
0	0.92	0.80	0.86	15
1	0.82	0.93	0.87	15
accuracy			0.87	30
macro avg	0.87	0.87	0.87	30
weighted avg	0.87	0.87	0.87	30
	precision	recall	f1-score	support
0	0.79	0.73	0.76	15
1	0.75	0.80	0.77	15
accuracy			0.77	30
macro avg	0.77	0.77	0.77	30
weighted avg	0.77	0.77	0.77	30
	precision	recall	f1-score	support
0	0.57	0.80	0.67	15
1	0.67	0.40	0.50	15
accuracy			0.60	30
macro avg	0.62	0.60	0.58	30
weighted avg	0.62	0.60	0.58	30

41) Binary Classification of Robust and Frail with MCI (76 samples) – 5-fold Cross Validation (SVM)

	precision	recall	f1-score	support
0	0.91	0.62	0.74	16
1	0.70	0.93	0.80	15
accuracy			0.77	31
macro avg	0.80	0.78	0.77	31
weighted avg	0.81	0.77	0.77	31

	precision	recall	f1-score	support
0	0.67	0.67	0.67	15
1	0.69	0.69	0.69	16
accuracy			0.68	31
macro avg	0.68	0.68	0.68	31
weighted avg	0.68	0.68	0.68	31
	precision	recall	f1-score	support
0	0.76	0.87	0.81	15
1	0.85	0.73	0.79	15
accuracy			0.80	30
macro avg	0.81	0.80	0.80	30
weighted avg	0.81	0.80	0.80	30
	precision	recall	f1-score	support
0	0.71	0.67	0.69	15
1	0.69	0.73	0.71	15
accuracy			0.70	30
macro avg	0.70	0.70	0.70	30
weighted avg	0.70	0.70	0.70	30
	precision	recall	f1-score	support
0	0.67	0.67	0.67	15
1	0.67	0.67	0.67	15
accuracy			0.67	30
macro avg	0.67	0.67	0.67	30
weighted avg	0.67	0.67	0.67	30

42) Binary Classification of Robust and Frail with MCI (76 samples) – 5-fold Cross Validation (RFC)

	precision	recall	f1-score	support
0	0.83	0.62	0.71	16
1	0.68	0.87	0.76	15
accuracy			0.74	31
macro avg	0.76	0.75	0.74	31
weighted avg	0.76	0.74	0.74	31
	precision	recall	f1-score	support
0	0.69	0.73	0.71	15
1	0.73	0.69	0.71	16
accuracy			0.71	31
macro avg	0.71	0.71	0.71	31
weighted avg	0.71	0.71	0.71	31
	precision	recall	f1-score	support
0	0.72	0.87	0.79	15
1	0.83	0.67	0.74	15
accuracy			0.77	30
macro avg	0.78	0.77	0.76	30
weighted avg	0.78	0.77	0.76	30
	precision	recall	f1-score	support
0	0.67	0.67	0.67	15
1	0.67	0.67	0.67	15
accuracy			0.67	30
macro avg	0.67	0.67	0.67	30
weighted avg	0.67	0.67	0.67	30
	precision	recall	f1-score	support
0	0.65	0.73	0.69	15
1	0.69	0.60	0.64	15
accuracy			0.67	30
macro avg	0.67	0.67	0.67	30
weighted avg	0.67	0.67	0.67	30

43) Binary Classification of Robust and Frail with MCI (343 samples) – Holdout  
Method (LOG)

```
Performance Metrics for Logistic Regression:  
0.77818  
[[ 94  37]  
 [ 24 120]]  


|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.80      | 0.72   | 0.76     | 131     |
| 1            | 0.76      | 0.83   | 0.80     | 144     |
| accuracy     |           |        | 0.78     | 275     |
| macro avg    | 0.78      | 0.78   | 0.78     | 275     |
| weighted avg | 0.78      | 0.78   | 0.78     | 275     |


```

44) Binary Classification of Robust and Frail with MCI (343 samples) – Holdout  
Method (LDA)

```
Performance Metrics for LDA:  
0.82182  
[[100  31]  
 [ 18 126]]  


|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.85      | 0.76   | 0.80     | 131     |
| 1            | 0.80      | 0.88   | 0.84     | 144     |
| accuracy     |           |        | 0.82     | 275     |
| macro avg    | 0.83      | 0.82   | 0.82     | 275     |
| weighted avg | 0.82      | 0.82   | 0.82     | 275     |


```

45) Binary Classification of Robust and Frail with MCI (343 samples) – Holdout

Method (KNN)

Performance Metrics for KNN:

0.75636

[[ 71 60]  
 [ 7 137]]

	precision	recall	f1-score	support
0	0.91	0.54	0.68	131
1	0.70	0.95	0.80	144
accuracy			0.76	275
macro avg	0.80	0.75	0.74	275
weighted avg	0.80	0.76	0.74	275

46) Binary Classification of Robust and Frail with MCI (343 samples) – Holdout

Method (CART)

Performance Metrics for CART:

0.77091

[[100 31]  
 [ 32 112]]

	precision	recall	f1-score	support
0	0.76	0.76	0.76	131
1	0.78	0.78	0.78	144
accuracy			0.77	275
macro avg	0.77	0.77	0.77	275
weighted avg	0.77	0.77	0.77	275

47) Binary Classification of Robust and Frail with MCI (343 samples) – Holdout

Method (GNB)

Performance Metrics for GNB:

0.77818

[[ 94 37]  
[ 24 120]]

	precision	recall	f1-score	support
0	0.80	0.72	0.76	131
1	0.76	0.83	0.80	144
accuracy			0.78	275
macro avg	0.78	0.78	0.78	275
weighted avg	0.78	0.78	0.78	275

48) Binary Classification of Robust and Frail with MCI (343 samples) – Holdout

Method (SVM)

Performance Metrics for SVM:

0.79273

[[ 94 37]  
[ 20 124]]

	precision	recall	f1-score	support
0	0.82	0.72	0.77	131
1	0.77	0.86	0.81	144
accuracy			0.79	275
macro avg	0.80	0.79	0.79	275
weighted avg	0.80	0.79	0.79	275

49) Binary Classification of Robust and Frail with MCI (343 samples) – Holdout Method (RFC)

```
Performance Metrics for RFC:
```

```
0.90182
```

```
[[117 14]
 [ 13 131]]
```

	precision	recall	f1-score	support
0	0.90	0.89	0.90	131
1	0.90	0.91	0.91	144
accuracy			0.90	275
macro avg	0.90	0.90	0.90	275
weighted avg	0.90	0.90	0.90	275

50) Binary Classification of Robust and Frail with MCI (343 samples) – 5-fold Cross Validation (LOG)

	precision	recall	f1-score	support
0	0.79	0.78	0.79	69
1	0.79	0.80	0.79	69
accuracy			0.79	138
macro avg	0.79	0.79	0.79	138
weighted avg	0.79	0.79	0.79	138
	precision	recall	f1-score	support
0	0.84	0.76	0.80	68
1	0.79	0.86	0.82	69
accuracy			0.81	137
macro avg	0.81	0.81	0.81	137
weighted avg	0.81	0.81	0.81	137

	precision	recall	f1-score	support
0	0.82	0.68	0.74	68
1	0.73	0.86	0.79	69
accuracy			0.77	137
macro avg	0.77	0.77	0.76	137
weighted avg	0.77	0.77	0.76	137
	precision	recall	f1-score	support
0	0.86	0.86	0.86	69
1	0.85	0.85	0.85	68
accuracy			0.85	137
macro avg	0.85	0.85	0.85	137
weighted avg	0.85	0.85	0.85	137
	precision	recall	f1-score	support
0	0.87	0.70	0.77	69
1	0.74	0.90	0.81	68
accuracy			0.80	137
macro avg	0.81	0.80	0.79	137
weighted avg	0.81	0.80	0.79	137

51) Binary Classification of Robust and Frail with MCI (343 samples) – 5-fold Cross Validation (LDA)

	precision	recall	f1-score	support
0	0.76	0.74	0.75	69
1	0.75	0.77	0.76	69
accuracy			0.75	138
macro avg	0.75	0.75	0.75	138
weighted avg	0.75	0.75	0.75	138
	precision	recall	f1-score	support
0	0.86	0.74	0.79	68
1	0.77	0.88	0.82	69
accuracy			0.81	137
macro avg	0.82	0.81	0.81	137
weighted avg	0.82	0.81	0.81	137

	precision	recall	f1-score	support
0	0.85	0.65	0.73	68
1	0.72	0.88	0.79	69
accuracy			0.77	137
macro avg	0.78	0.77	0.76	137
weighted avg	0.78	0.77	0.76	137
	precision	recall	f1-score	support
0	0.82	0.86	0.84	69
1	0.85	0.81	0.83	68
accuracy			0.83	137
macro avg	0.83	0.83	0.83	137
weighted avg	0.83	0.83	0.83	137
	precision	recall	f1-score	support
0	0.86	0.72	0.79	69
1	0.76	0.88	0.82	68
accuracy			0.80	137
macro avg	0.81	0.80	0.80	137
weighted avg	0.81	0.80	0.80	137

52) Binary Classification of Robust and Frail with MCI (343 samples) – 5-fold Cross Validation (KNN)

	precision	recall	f1-score	support
0	0.92	0.68	0.78	69
1	0.75	0.94	0.83	69
accuracy			0.81	138
macro avg	0.83	0.81	0.81	138
weighted avg	0.83	0.81	0.81	138
	precision	recall	f1-score	support
0	0.91	0.62	0.74	68
1	0.71	0.94	0.81	69
accuracy			0.78	137
macro avg	0.81	0.78	0.77	137
weighted avg	0.81	0.78	0.77	137

	precision	recall	f1-score	support
0	1.00	0.44	0.61	68
1	0.64	1.00	0.78	69
accuracy			0.72	137
macro avg	0.82	0.72	0.70	137
weighted avg	0.82	0.72	0.70	137
	precision	recall	f1-score	support
0	0.98	0.62	0.76	69
1	0.72	0.99	0.83	68
accuracy			0.80	137
macro avg	0.85	0.80	0.80	137
weighted avg	0.85	0.80	0.80	137
	precision	recall	f1-score	support
0	1.00	0.49	0.66	69
1	0.66	1.00	0.80	68
accuracy			0.74	137
macro avg	0.83	0.75	0.73	137
weighted avg	0.83	0.74	0.73	137

53) Binary Classification of Robust and Frail with MCI (343 samples) – 5-fold Cross Validation (CART)

	precision	recall	f1-score	support
0	0.76	0.86	0.80	69
1	0.83	0.72	0.78	69
accuracy			0.79	138
macro avg	0.79	0.79	0.79	138
weighted avg	0.79	0.79	0.79	138
	precision	recall	f1-score	support
0	0.81	0.82	0.82	68
1	0.82	0.81	0.82	69
accuracy			0.82	137
macro avg	0.82	0.82	0.82	137
weighted avg	0.82	0.82	0.82	137

	precision	recall	f1-score	support
0	0.86	0.75	0.80	68
1	0.78	0.88	0.83	69
accuracy			0.82	137
macro avg	0.82	0.82	0.82	137
weighted avg	0.82	0.82	0.82	137
	precision	recall	f1-score	support
0	0.88	0.67	0.76	69
1	0.73	0.91	0.81	68
accuracy			0.79	137
macro avg	0.81	0.79	0.79	137
weighted avg	0.81	0.79	0.79	137
	precision	recall	f1-score	support
0	0.89	0.72	0.80	69
1	0.77	0.91	0.83	68
accuracy			0.82	137
macro avg	0.83	0.82	0.82	137
weighted avg	0.83	0.82	0.82	137

54) Binary Classification of Robust and Frail with MCI (343 samples) – 5-fold Cross Validation (GNB)

	precision	recall	f1-score	support
0	0.70	0.84	0.76	69
1	0.80	0.64	0.71	69
accuracy			0.74	138
macro avg	0.75	0.74	0.74	138
weighted avg	0.75	0.74	0.74	138
	precision	recall	f1-score	support
0	0.74	0.72	0.73	68
1	0.73	0.75	0.74	69
accuracy			0.74	137
macro avg	0.74	0.74	0.74	137
weighted avg	0.74	0.74	0.74	137

	precision	recall	f1-score	support
0	0.71	0.72	0.72	68
1	0.72	0.71	0.72	69
accuracy			0.72	137
macro avg	0.72	0.72	0.72	137
weighted avg	0.72	0.72	0.72	137
	precision	recall	f1-score	support
0	0.77	0.86	0.81	69
1	0.83	0.74	0.78	68
accuracy			0.80	137
macro avg	0.80	0.80	0.79	137
weighted avg	0.80	0.80	0.79	137
	precision	recall	f1-score	support
0	0.68	0.80	0.73	69
1	0.75	0.62	0.68	68
accuracy			0.71	137
macro avg	0.71	0.71	0.71	137
weighted avg	0.71	0.71	0.71	137

55) Binary Classification of Robust and Frail with MCI (343 samples) – 5-fold Cross Validation (SVM)

	precision	recall	f1-score	support
0	0.81	0.72	0.76	69
1	0.75	0.83	0.79	69
accuracy			0.78	138
macro avg	0.78	0.78	0.77	138
weighted avg	0.78	0.78	0.77	138
	precision	recall	f1-score	support
0	0.84	0.71	0.77	68
1	0.75	0.87	0.81	69
accuracy			0.79	137
macro avg	0.80	0.79	0.79	137
weighted avg	0.80	0.79	0.79	137

	precision	recall	f1-score	support
0	0.87	0.60	0.71	68
1	0.70	0.91	0.79	69
accuracy			0.76	137
macro avg	0.79	0.76	0.75	137
weighted avg	0.79	0.76	0.75	137
	precision	recall	f1-score	support
0	0.89	0.84	0.87	69
1	0.85	0.90	0.87	68
accuracy			0.87	137
macro avg	0.87	0.87	0.87	137
weighted avg	0.87	0.87	0.87	137
	precision	recall	f1-score	support
0	0.88	0.65	0.75	69
1	0.72	0.91	0.81	68
accuracy			0.78	137
macro avg	0.80	0.78	0.78	137
weighted avg	0.80	0.78	0.78	137

56) Binary Classification of Robust and Frail with MCI (343 samples) – 5-fold Cross Validation (RFC)

	precision	recall	f1-score	support
0	0.86	0.99	0.92	69
1	0.98	0.84	0.91	69
accuracy			0.91	138
macro avg	0.92	0.91	0.91	138
weighted avg	0.92	0.91	0.91	138
	precision	recall	f1-score	support
0	0.94	0.91	0.93	68
1	0.92	0.94	0.93	69
accuracy			0.93	137
macro avg	0.93	0.93	0.93	137
weighted avg	0.93	0.93	0.93	137

	precision	recall	f1-score	support
0	0.95	0.84	0.89	68
1	0.86	0.96	0.90	69
accuracy			0.90	137
macro avg	0.90	0.90	0.90	137
weighted avg	0.90	0.90	0.90	137
	precision	recall	f1-score	support
0	0.98	0.93	0.96	69
1	0.93	0.99	0.96	68
accuracy			0.96	137
macro avg	0.96	0.96	0.96	137
weighted avg	0.96	0.96	0.96	137
	precision	recall	f1-score	support
0	1.00	0.86	0.92	69
1	0.87	1.00	0.93	68
accuracy			0.93	137
macro avg	0.94	0.93	0.93	137
weighted avg	0.94	0.93	0.93	137

57) Binary Classification of Robust and Frail with MCI (100 samples) – Holdout Method (LOG)

Performance Metrics for Logistic Regression:

0.8

[[32 10]  
[ 6 32]]

	precision	recall	f1-score	support
0	0.84	0.76	0.80	42
1	0.76	0.84	0.80	38
accuracy			0.80	80
macro avg	0.80	0.80	0.80	80
weighted avg	0.80	0.80	0.80	80

58) Binary Classification of Robust and Frail with MCI (100 samples) – Holdout

Method (LDA)

Performance Metrics for LDA:

0.7375

[[30 12]  
[ 9 29]]

	precision	recall	f1-score	support
0	0.77	0.71	0.74	42
1	0.71	0.76	0.73	38
accuracy			0.74	80
macro avg	0.74	0.74	0.74	80
weighted avg	0.74	0.74	0.74	80

59) Binary Classification of Robust and Frail with MCI (100 samples) – Holdout

Method (KNN)

Performance Metrics for KNN:

0.725

[[36 6]  
[16 22]]

	precision	recall	f1-score	support
0	0.69	0.86	0.77	42
1	0.79	0.58	0.67	38
accuracy			0.73	80
macro avg	0.74	0.72	0.72	80
weighted avg	0.74	0.72	0.72	80

60) Binary Classification of Robust and Frail with MCI (100 samples) – Holdout

Method (CART)

Performance Metrics for CART:

0.7625

[[26 16]  
[ 3 35]]

	precision	recall	f1-score	support
0	0.90	0.62	0.73	42
1	0.69	0.92	0.79	38
accuracy			0.76	80
macro avg	0.79	0.77	0.76	80
weighted avg	0.80	0.76	0.76	80

61) Binary Classification of Robust and Frail with MCI (100 samples) – Holdout

Method (GNB)

Performance Metrics for GNB:

0.8

[[32 10]  
[ 6 32]]

	precision	recall	f1-score	support
0	0.84	0.76	0.80	42
1	0.76	0.84	0.80	38
accuracy			0.80	80
macro avg	0.80	0.80	0.80	80
weighted avg	0.80	0.80	0.80	80

62) Binary Classification of Robust and Frail with MCI (100 samples) – Holdout

Method (SVM)

Performance Metrics for SVM:

0.8

[[32 10]  
[ 6 32]]

	precision	recall	f1-score	support
0	0.84	0.76	0.80	42
1	0.76	0.84	0.80	38
accuracy			0.80	80
macro avg	0.80	0.80	0.80	80
weighted avg	0.80	0.80	0.80	80

63) Binary Classification of Robust and Frail with MCI (100 samples) – Holdout

Method (RFC)

Performance Metrics for RFC:

0.825

[[32 10]  
[ 4 34]]

	precision	recall	f1-score	support
0	0.89	0.76	0.82	42
1	0.77	0.89	0.83	38
accuracy			0.82	80
macro avg	0.83	0.83	0.82	80
weighted avg	0.83	0.82	0.82	80

64) Binary Classification of Robust and Frail with MCI (100 samples) – 5-fold Cross Validation (LOG)

	precision	recall	f1-score	support
0	0.71	0.75	0.73	20
1	0.74	0.70	0.72	20
accuracy			0.73	40
macro avg	0.73	0.72	0.72	40
weighted avg	0.73	0.72	0.72	40
	precision	recall	f1-score	support
0	0.86	0.60	0.71	20
1	0.69	0.90	0.78	20
accuracy			0.75	40
macro avg	0.77	0.75	0.74	40
weighted avg	0.77	0.75	0.74	40
	precision	recall	f1-score	support
0	0.73	0.80	0.76	20
1	0.78	0.70	0.74	20
accuracy			0.75	40
macro avg	0.75	0.75	0.75	40
weighted avg	0.75	0.75	0.75	40
	precision	recall	f1-score	support
0	0.78	0.70	0.74	20
1	0.73	0.80	0.76	20
accuracy			0.75	40
macro avg	0.75	0.75	0.75	40
weighted avg	0.75	0.75	0.75	40
	precision	recall	f1-score	support
0	0.95	0.90	0.92	20
1	0.90	0.95	0.93	20
accuracy			0.93	40
macro avg	0.93	0.93	0.92	40
weighted avg	0.93	0.93	0.92	40

65) Binary Classification of Robust and Frail with MCI (100 samples) – 5-fold Cross Validation (LDA)

	precision	recall	f1-score	support
0	0.70	0.70	0.70	20
1	0.70	0.70	0.70	20
accuracy			0.70	40
macro avg	0.70	0.70	0.70	40
weighted avg	0.70	0.70	0.70	40
	precision	recall	f1-score	support
0	0.92	0.55	0.69	20
1	0.68	0.95	0.79	20
accuracy			0.75	40
macro avg	0.80	0.75	0.74	40
weighted avg	0.80	0.75	0.74	40
	precision	recall	f1-score	support
0	0.75	0.75	0.75	20
1	0.75	0.75	0.75	20
accuracy			0.75	40
macro avg	0.75	0.75	0.75	40
weighted avg	0.75	0.75	0.75	40
	precision	recall	f1-score	support
0	0.71	0.75	0.73	20
1	0.74	0.70	0.72	20
accuracy			0.73	40
macro avg	0.73	0.72	0.72	40
weighted avg	0.73	0.72	0.72	40
	precision	recall	f1-score	support
0	1.00	0.90	0.95	20
1	0.91	1.00	0.95	20
accuracy			0.95	40
macro avg	0.95	0.95	0.95	40
weighted avg	0.95	0.95	0.95	40

66) Binary Classification of Robust and Frail with MCI (100 samples) – 5-fold Cross Validation (KNN)

	precision	recall	f1-score	support
0	0.66	0.95	0.78	20
1	0.91	0.50	0.65	20
accuracy			0.73	40
macro avg	0.78	0.72	0.71	40
weighted avg	0.78	0.72	0.71	40
	precision	recall	f1-score	support
0	0.58	0.70	0.64	20
1	0.62	0.50	0.56	20
accuracy			0.60	40
macro avg	0.60	0.60	0.60	40
weighted avg	0.60	0.60	0.60	40
	precision	recall	f1-score	support
0	0.70	0.95	0.81	20
1	0.92	0.60	0.73	20
accuracy			0.78	40
macro avg	0.81	0.77	0.77	40
weighted avg	0.81	0.78	0.77	40
	precision	recall	f1-score	support
0	0.65	0.65	0.65	20
1	0.65	0.65	0.65	20
accuracy			0.65	40
macro avg	0.65	0.65	0.65	40
weighted avg	0.65	0.65	0.65	40
	precision	recall	f1-score	support
0	0.86	0.95	0.90	20
1	0.94	0.85	0.89	20
accuracy			0.90	40
macro avg	0.90	0.90	0.90	40
weighted avg	0.90	0.90	0.90	40

67) Binary Classification of Robust and Frail with MCI (100 samples) – 5-fold Cross Validation (CART)

	precision	recall	f1-score	support
0	0.60	0.60	0.60	20
1	0.60	0.60	0.60	20
accuracy			0.60	40
macro avg	0.60	0.60	0.60	40
weighted avg	0.60	0.60	0.60	40
	precision	recall	f1-score	support
0	0.75	0.75	0.75	20
1	0.75	0.75	0.75	20
accuracy			0.75	40
macro avg	0.75	0.75	0.75	40
weighted avg	0.75	0.75	0.75	40
	precision	recall	f1-score	support
0	0.83	0.75	0.79	20
1	0.77	0.85	0.81	20
accuracy			0.80	40
macro avg	0.80	0.80	0.80	40
weighted avg	0.80	0.80	0.80	40
	precision	recall	f1-score	support
0	0.67	0.70	0.68	20
1	0.68	0.65	0.67	20
accuracy			0.68	40
macro avg	0.68	0.68	0.67	40
weighted avg	0.68	0.68	0.67	40
	precision	recall	f1-score	support
0	0.91	1.00	0.95	20
1	1.00	0.90	0.95	20
accuracy			0.95	40
macro avg	0.95	0.95	0.95	40
weighted avg	0.95	0.95	0.95	40

68) Binary Classification of Robust and Frail with MCI (100 samples) – 5-fold Cross Validation (GNB)

	precision	recall	f1-score	support
0	0.74	0.85	0.79	20
1	0.82	0.70	0.76	20
accuracy			0.78	40
macro avg	0.78	0.77	0.77	40
weighted avg	0.78	0.78	0.77	40
	precision	recall	f1-score	support
0	0.84	0.80	0.82	20
1	0.81	0.85	0.83	20
accuracy			0.82	40
macro avg	0.83	0.82	0.82	40
weighted avg	0.83	0.82	0.82	40
	precision	recall	f1-score	support
0	0.68	0.85	0.76	20
1	0.80	0.60	0.69	20
accuracy			0.73	40
macro avg	0.74	0.72	0.72	40
weighted avg	0.74	0.72	0.72	40
	precision	recall	f1-score	support
0	0.65	0.85	0.74	20
1	0.79	0.55	0.65	20
accuracy			0.70	40
macro avg	0.72	0.70	0.69	40
weighted avg	0.72	0.70	0.69	40
	precision	recall	f1-score	support
0	0.62	1.00	0.77	20
1	1.00	0.40	0.57	20
accuracy			0.70	40
macro avg	0.81	0.70	0.67	40
weighted avg	0.81	0.70	0.67	40

69) Binary Classification of Robust and Frail with MCI (100 samples) – 5-fold Cross Validation (SVM)

	precision	recall	f1-score	support
0	0.68	0.75	0.71	20
1	0.72	0.65	0.68	20
accuracy			0.70	40
macro avg	0.70	0.70	0.70	40
weighted avg	0.70	0.70	0.70	40
	precision	recall	f1-score	support
0	0.88	0.70	0.78	20
1	0.75	0.90	0.82	20
accuracy			0.80	40
macro avg	0.81	0.80	0.80	40
weighted avg	0.81	0.80	0.80	40
	precision	recall	f1-score	support
0	0.74	0.85	0.79	20
1	0.82	0.70	0.76	20
accuracy			0.78	40
macro avg	0.78	0.77	0.77	40
weighted avg	0.78	0.78	0.77	40
	precision	recall	f1-score	support
0	0.78	0.70	0.74	20
1	0.73	0.80	0.76	20
accuracy			0.75	40
macro avg	0.75	0.75	0.75	40
weighted avg	0.75	0.75	0.75	40
	precision	recall	f1-score	support
0	0.95	0.95	0.95	20
1	0.95	0.95	0.95	20
accuracy			0.95	40
macro avg	0.95	0.95	0.95	40
weighted avg	0.95	0.95	0.95	40

70) Binary Classification of Robust and Frail with MCI (100 samples) – 5-fold Cross Validation (RFC)

	precision	recall	f1-score	support
0	0.76	0.80	0.78	20
1	0.79	0.75	0.77	20
accuracy			0.78	40
macro avg	0.78	0.78	0.77	40
weighted avg	0.78	0.78	0.77	40
	precision	recall	f1-score	support
0	0.78	0.70	0.74	20
1	0.73	0.80	0.76	20
accuracy			0.75	40
macro avg	0.75	0.75	0.75	40
weighted avg	0.75	0.75	0.75	40
	precision	recall	f1-score	support
0	0.94	0.85	0.89	20
1	0.86	0.95	0.90	20
accuracy			0.90	40
macro avg	0.90	0.90	0.90	40
weighted avg	0.90	0.90	0.90	40
	precision	recall	f1-score	support
0	0.84	0.80	0.82	20
1	0.81	0.85	0.83	20
accuracy			0.82	40
macro avg	0.83	0.82	0.82	40
weighted avg	0.83	0.82	0.82	40
	precision	recall	f1-score	support
0	1.00	1.00	1.00	20
1	1.00	1.00	1.00	20
accuracy			1.00	40
macro avg	1.00	1.00	1.00	40
weighted avg	1.00	1.00	1.00	40