FYP 1 Draft Interim Report

Classification of Cognitive Frailty (CF) in Elderly People using Features Extracted from Blood Samples

by

Shahidan Bin Idris
25204
Associate Professor Dr Nasreen binti Badruddin

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 proposes a model to classify patients into different levels of CF, using blood-based parameters from blood samples. The model will also be used to identify the top performing biomarkers that are associated with CF. In the study, the dataset from the Malaysian Elders Longitudinal Research (MELoR) is used, which consists of socio-demographic and medical data of older adults. The dataset is cleaned from errors and missing data, followed by pre-processing and input optimisation. 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) are trained. From the preliminary results, Logistic Regression had the highest holdout method accuracy of 38.2%, followed by RFC at 37.7%. For cross validation, Logistic Regression maintained with the highest accuracy of 37% with a standard deviation of 0.04, followed by LDA with an accuracy of 36% with a standard deviation of 0.03. Further analysis is done to identify the significant class imbalance in the dataset, proposing 3 methods of handling the imbalance – Binary Classification (Robust and Non-Robust), Binary Classification (Robust and Frail with MCI), and SMOTE. Using binary classification (2 classes) instead of 6 classes produced better accuracy overall, with SMOTE (using 6 classes) producing significant improvement in accuracy. While the performance for the models are improved using these approaches, some important concerns and drawbacks were identified and brought up for each approach, providing key takeaways and lessons for future research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND OF STUDY

The aging population worldwide is increasing over the years. This trend also holds true for Malaysia, as shown in Figure 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" [2].

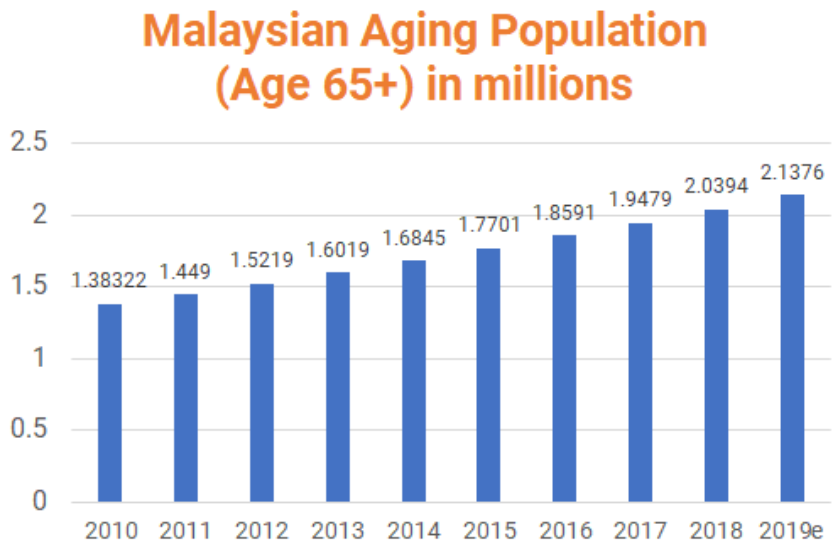


FIGURE 1. Population by age group, sex and ethnic group, 2010 - 2019e, Malaysia [1]

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 a standard clinical diagnosis tool for CF.

1.2 PROBLEM STATEMENT

"There is no existing classification model to predict Cognitive Frailty in elderly people using blood samples."

- Existing models classify Cognitive Frailty based on questionnaire responses and physical/cognitive assessment results, rather than physiological parameters (e.g. blood-based parameters)
- Models predicting the condition using physiological parameters have been focused on physical frailty and mild cognitive impairment (MCI) separately
- Attempts made have mostly conducted clustering (instead of classification) or used statistical analysis techniques rather than modern classification algorithms

1.3 OBJECTIVE 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

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 made 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 defined 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 is 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 are 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 with 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. 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 [17], [18], while physical frailty symptoms have been found to manifest before the onset of dementia [19], [20]. In 2013, the International Academy on Nutrition and Aging (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 for 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 is done according to weakness, slowness, low level of physical activity, exhaustion and poor endurance, and weight loss. The table below 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 1.

TABLE 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, >= 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 2. 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 6 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 aging process closely relates with an acceleration in inflammatory activity in the body [21]. 3 of the inflammatory markers that are most researched are the Interleukin 6 (IL-6), C-Reactive Protein (CRP), and Tumor Necrosis Factor Alpha (TNFα). IL-6 and TNFα 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 [22], [23]. 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 TNFα, CRP is strongly related to the frailty syndrome [24], [25]. 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 [26 - 28]. 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 [29 - 32]. 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) [33]. 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 β-amyloid peptides (1-42), Tau levels, and amyloid-Tau ratio [34], [35]. 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 the blood plasma has been shown to be altered when there is neurodegeneration [36]. Higher levels of clusterin has been found in patients with MCI compared with patients with normal cognitive function. High levels of clusterin has 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 [37]. 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 [38]. 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, has 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 [39].

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 did 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 [40]. 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 [41]. 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 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 ± 0.70% accuracy, while the Random Forest model achieved 77.46 ± 0.45% accuracy [42]. 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 biomarkers, such as blood.

CHAPTER 3

METHODOLOGY

3.1 RESEARCH METHODOLOGY

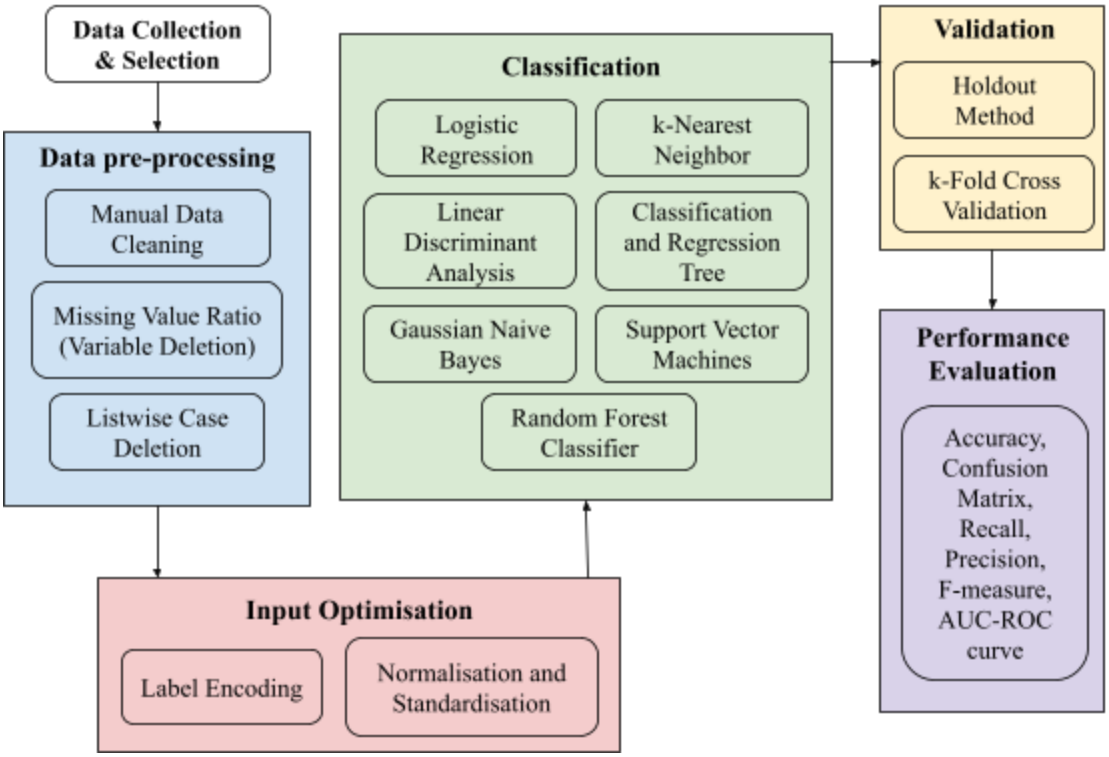


FIGURE 2. Research Methodology Flow Diagram

The figure above shows the research methodology for the project. The work for the project is separated into 6 parts, which are the data collection & selection, data pre-processing, input optimisation, classification, validation, and performance evaluation. Each step is explained as follows:

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. [43] 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) [44]. 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 [45], [46].

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 3. 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 [47]. 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 10^9) 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 '>54' 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 [48]. 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 [49]. 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 4. 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 47 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.

Optimisation of the dataset values is important as machine learning models rely on good data to perform well [50]. 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, σ, 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 [51 - 53]. 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 Algorithms

7 different machine learning algorithms are used to classify the participants into one of 6 CF groups 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 between the 7 models and the best performing model is selected as the final classification model for the project.

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 are 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.

3.2 TOOLS

1. Python (Anaconda Distribution)

2. Jupyter Notebook
3. Scikit-learn

3.3 PROJECT MILESTONES AND GANTT CHART

TABLE 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. 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 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 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 PRELIMINARY RESULTS

Using the research methodology highlighted previously, 7 algorithms were used to train 7 models to classify CF into 6 groups. The tables below show the performance evaluation for each algorithm for both holdout method and k-fold cross validation. The confusion matrices and performance evaluation per class for holdout method is provided in Appendix 5, while the performance evaluation per class for each iteration of k-fold cross validation is provided in Appendix 6.

TABLE 6. Holdout Method Performance Evaluation

Algorithms	Accuracy	Precision	Recall	F-Measure
Logistic Regression (Log)	38.2%	17%	21%	18%
Linear Discrimant Analysis (LDA)	35.7%	23%	21%	20%
K-Nearest Neighbor (kNN)	33.5%	14%	18%	15%
Classification and Regression Tree (CART)	25.1%	20%	19%	19%
Gaussian Naive Bayes (GNB)	36.9%	17%	21%	18%
Support Vector Machines (SVM)	36.2%	6%	17%	9%
Random Forest Classifier (RFC)	37.7%	16%	21%	18%

TABLE 7. 5-fold Cross Validation Performance Evaluation

Algorithms	Accuracy	Precision	Recall	F-Measure
Logistic Regression (Log)	37% (std. dev. 0.04)	19.6%	21.0%	17.4%
Linear Discrimant Analysis (LDA)	36% (std. dev. 0.03)	21.2%	21.4%	19.8%
K-Nearest Neighbor (kNN)	31% (std. dev. 0.02)	18.2%	17.8%	16.2%
Classification and Regression Tree (CART)	26% (std. dev. 0.04)	20.2%	19.6%	19.6%
Gaussian Naive Bayes (GNB)	33% (std. dev. 0.03)	22.2%	22.0%	20.6%
Support Vector Machines (SVM)	34% (std. dev. 0.00)	6.0%	17.0%	8.0%
Random Forest Classifier (RFC)	35% (std. dev. 0.03)	20.4%	19.4%	16.0%

For the holdout method, Logistic Regression had the highest accuracy of 38.2%, followed by RFC at 37.7%. For cross validation, Logistic Regression maintained with the highest accuracy of 37% with a standard deviation of 0.04, followed by LDA with an accuracy of 36% with a standard deviation of 0.03.

4.2 DISCUSSION

While the accuracy scores for all models were higher than the probability of classifying a sample into one out of the 6 classes (⅙ = 16.67%), the accuracy score achieved from these results were unsatisfactory. After further analysing the evaluation scores, confusion matrices and class sizes, it was found that the dataset faces a significant class imbalance problem. A common mitigation approach is to do undersampling and oversampling to balance the class samples. It was suggested to carry out the approach through the following possible methods:

4.2.1 Binary Classification (Robust and Non-Robust)

The 5 classes that are not 'Robust' were grouped as 'Non-Robust'. This created a new class size of Robust with 343 samples and Non-Robust with 672 samples. From the majority class (non-robust), 343 instances were randomly sampled to create an even balance between the two classes. The same research methodology was carried out and the performance evaluation is as follows:

TABLE 8. Robust vs Non-Robust Holdout Method Accuracy Score

Algorithms	Accuracy
Logistic Regression (Log)	56.4%
Linear Discrimant Analysis (LDA)	60.4%
K-Nearest Neighbor (kNN)	53.8%
Classification and Regression Tree (CART)	51.6%
Gaussian Naive Bayes (GNB)	57.1%
Support Vector Machines (SVM)	60.0%
Random Forest Classifier (RFC)	59.6%

TABLE 9. Robust vs Non-Robust 5-fold Cross Validation Accuracy Score

Algorithms	Accuracy
Logistic Regression (Log)	62% (std. dev. 0.04)
Linear Discrimant Analysis (LDA)	64% (std. dev. 0.03)
K-Nearest Neighbor (kNN)	54% (std. dev. 0.04)
Classification and Regression Tree (CART)	55% (std. dev. 0.04)

Gaussian Naive Bayes (GNB)	56% (std. dev. 0.02)
Support Vector Machines (SVM)	63% (std. dev. 0.03)
Random Forest Classifier (RFC)	65% (std. dev. 0.05)

The accuracy scores were shown to increase significantly compared to using 6 classes. However, the accuracy was only slightly higher than the probability of classifying into 2 classes (½ = 50.0%). Furthermore, the classification of patients into Robust and Non-Robust does not clearly achieve the objectives of the project, which is to classify CF in patients. Instead, the classification of Robust and Non-Robust merely distinguishes between healthy and unhealthy individuals.

4.2.2 Binary Classification (Robust and Frail + MCI)

Only 2 out of 6 classes were used for the classification (Robust and Frail + MCI). This method used the extreme ends of the CF spectrum (most healthy vs most unhealthy) as the dataset. The class samples were Robust with 343 samples and Frail + MCI with 76 samples. Same as before, Robust was randomly undersampled for 76 samples, matching Frail + MCI. The table below shows the performance evaluation:

TABLE 10. Robust vs Frail + MCI Holdout Method Accuracy Score

Algorithms	Accuracy
Logistic Regression (Log)	60.7%
Linear Discrimant Analysis (LDA)	59.0%
K-Nearest Neighbor (kNN)	49.2%
Classification and Regression Tree (CART)	59%
Gaussian Naive Bayes (GNB)	70.5%

Support Vector Machines (SVM)	55.7%
Random Forest Classifier (RFC)	63.9%

TABLE 11. Robust vs Frail + MCI 5-fold Cross Validation Accuracy Score

Algorithms	Accuracy
Logistic Regression (Log)	72% (std. dev. 0.08)
Linear Discrimant Analysis (LDA)	69% (std. dev. 0.04)
K-Nearest Neighbor (kNN)	65% (std. dev. 0.05)
Classification and Regression Tree (CART)	60% (std. dev. 0.05)
Gaussian Naive Bayes (GNB)	74% (std. dev. 0.07)
Support Vector Machines (SVM)	73% (std. dev. 0.08)
Random Forest Classifier (RFC)	68% (std. dev. 0.10)

The accuracy scores greatly improved up to a satisfactory level. Compared to the previous method, the Robust vs Frail + MCI classification is better at achieving the project results – patients with both frailty and MCI can be distinguished from healthy individuals. However, the resulting dataset used for training and testing from this method is very small (total of 152 samples). The higher performance evaluation scores could possibly be attributed to a small test set, rather than good model performance.

4.2.3 Synthetic Minority Over-sampling Technique (SMOTE)

SMOTE is a data augmentation technique used to oversample the minority class in a dataset [12]. In this method, the remaining 5 classes are oversampled to match the number of samples in the majority class (343 samples for Robust). The performance evaluation for this method is shown:

TABLE 12.Holdout Method Accuracy Score using SMOTE

Algorithms	Accuracy
Logistic Regression (Log)	40.4%
Linear Discrimant Analysis (LDA)	42.1%
K-Nearest Neighbor (kNN)	56.8%
Classification and Regression Tree (CART)	52.3%
Gaussian Naive Bayes (GNB)	44.9%
Support Vector Machines (SVM)	41.3%
Random Forest Classifier (RFC)	72.8%

TABLE 13. 5-fold Cross Validation Accuracy Score using SMOTE

Algorithms	Accuracy
Logistic Regression (Log)	43% (std. dev. 0.08)
Linear Discrimant Analysis (LDA)	45% (std. dev. 0.04)
K-Nearest Neighbor (kNN)	63% (std. dev. 0.05)
Classification and Regression Tree (CART)	57% (std. dev. 0.05)
Gaussian Naive Bayes (GNB)	44% (std.

	dev. 0.07)
Support Vector Machines (SVM)	44% (std. dev. 0.08)
Random Forest Classifier (RFC)	76% (std. dev. 0.10)

As this classification is done for 6 classes, the accuracy scores for the models increased significantly, with the RFC model's accuracy increasing the most. While SMOTE is a powerful technique for oversampling, it is noted that using oversampling techniques to generate synthetic data does not translate to real-life model performance. While the accuracy for the RFC model may be high, the model may not perform as well when used on outside data. After further analysing the confusion matrices, it is found that RFC has predicted perfectly the smallest class (Frailty with originally 7 samples). This means that SMOTE synthesises data that is "easily predictable" for RFC. This is a concern that needs to be remembered when using SMOTE for this project.

CHAPTER 5
CONCLUSION & FUTURE WORK

5.1 CONCLUSION

The growing population worldwide comes hand-in-hand with the increasing population of the aging community. This increase in population brings about challenges for medical practitioners and researchers in handling age-related diseases, such as frailty and cognitive impairment. 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.

The preliminary study conducted covers the domain-specific topics, such as frailty, cognitive impairment, CF, and its diagnosis, while also analysing the potential biomarkers and current work that is being done to enable CF prediction. The literature review has specified the definitions employed for this project with regards to the disease components, while highlighting the scope of the study in terms of the choice of dataset and parameters selected.

The research methodology designed and proposed is inspired by the current practices and techniques used in the industry. While the proposed techniques and approaches are selected based on the best performance and outcomes according to preliminary studies, exploration and actual implementation may give better insights on a more suitable methodology for the project.

The preliminary results produced using the designed methodology serve as valuable progress for the project. While the results have shown weaknesses in the techniques applied, they have also shown equally vital information and key lessons to be used on improving the project. Majority of the time and effort for the project will be put into translating these key lessons into better results.

5.2 FUTURE WORK

As described in the results and discussion, there are key concerns with regards to the nature of the dataset, as well as the suitable techniques to be used on the dataset. A major challenge for the project is in handling the extreme class imbalance in the dataset. Much of the future work will be dedicated to exploring more data transformation and feature selection techniques which are hoped to improve the project results. Furthermore, it is considered to implement more advanced algorithms for feature extraction and classification, such as artificial neural networks. Further intensive research as well as trial-and-error will show its viability in helping with the classification tasks. This will also be an important future work for the project.

REFERENCES

[1] "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. [Accessed: 09-Mar-2021].

[2] 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.

[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. Albarede, 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] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] P. Soysal et al., "Inflammation and frailty in the elderly: A systematic review and meta-analysis," *Ageing Res. Rev.*, vol. 31, pp. 1–8, 2016.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] K. Blennow and H. Hampel, “CSF markers for incipient Alzheimer’s disease,” *Lancet Neurol.*, vol. 2, no. 10, pp. 605–613, 2003.

[36] 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.

[37] Z.-X. Yin et al., “High normal plasma triglycerides are associated with preserved cognitive function in Chinese oldest-old,” *Age Ageing*, vol. 41, no. 5, pp. 600–606, 2012.

[38] 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.

[39] 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.

[40] F. Babič, 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,

[41] 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.

[42] 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.

[43] *MELoR*. [Online]. Available: https://melor-um.weebly.com/. [Accessed: 06-Mar-2021].

[44] 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.

[45] 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.

[46] 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.

[47] J. I. Maletic and A. Marcus, "Data Cleansing," *Data Mining and Knowledge Discovery Handbook*, pp. 21–36.

[48] 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-learning-1a6326adf79e. [Accessed: 06-Mar-2021].

[49] J. Scheffer, "Dealing with Missing Data," *Research Letters in the Information and Mathematical Sciences*, vol. 3, 2002.

[50] I. H. Witten, E. Frank, and M. A. Hall, *Data mining: practical machine learning tools and techniques*. Amsterdam: Elsevier, 2011.

[51] K. Muralidharan, "A note on transformation, standardization and normalization," *International Journal of Operations and Quantitative Management*, vol. 9, 2010.

[52] P. J. M. Ali, "Data Normalization and Standardization: A Technical Report," *Machine Learning Technical Reports*, vol. 1, no. 1, 2014 .

[53] 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].

[54] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *arXiv [cs.AI]*, 2011

APPENDICES

Appendix 1

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 ≥ 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).

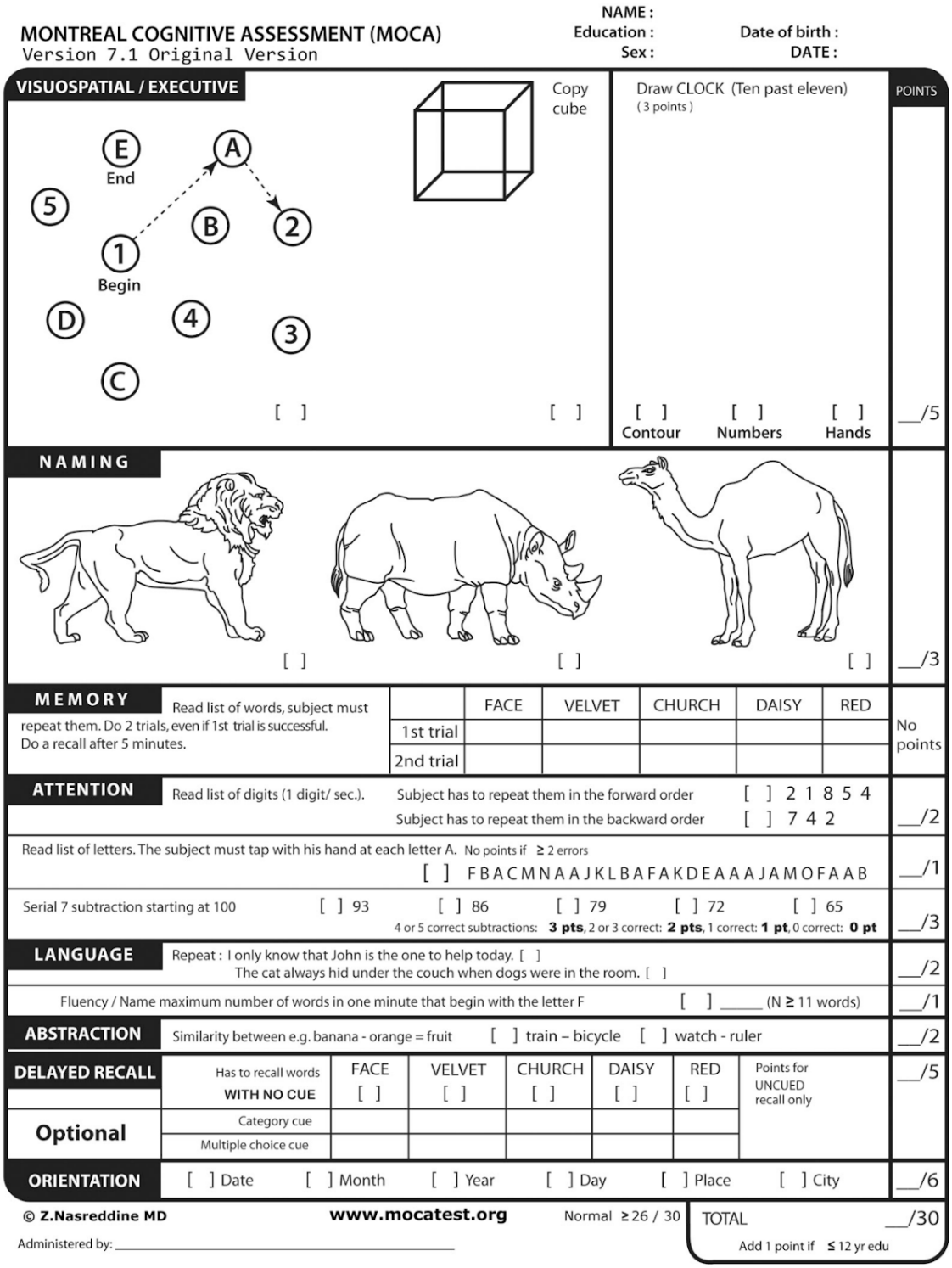
Men	*Cutoff for Time to Walk 15 feet criterion for frailty*
Height ≤ 173 cm	≥7 seconds
Height > 173 cm	≥6 seconds
Women	
Height ≤ 159 cm	≥7 seconds
Height > 159 cm	≥6 seconds

- **Grip Strength**, stratified by gender and body mass index (BMI) quartiles:

Men	*Cutoff for grip strength (Kg) criterion for frailty*
BMI ≤ 24	≤29
BMI 24.1–26	≤30
BMI 26.1–28	≤30
BMI > 28	≤32
Women	
BMI ≤ 23	≤17
BMI 23.1–26	≤17.3
BMI 26.1–29	≤18
BMI > 29	≤21

Appendix 2

MoCa Test (in English)



MONTREAL COGNITIVE ASSESSMENT (MOCA)
Version 7.1 Original Version

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

VISUOSPATIAL / EXECUTIVE — POINTS

Copy cube [ ]

Draw CLOCK (Ten past eleven) (3 points) [ ] [ ] [ ] /5
Contour Numbers Hands

NAMING [ ] [ ] [ ] /3

MEMORY Read list of words, subject must repeat them. Do 2 trials, even if 1st trial is successful. Do a recall after 5 minutes. — No points

	FACE	VELVET	CHURCH	DAISY	RED
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
Subject has to repeat them in the backward order [ ] 7 4 2 /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
4 or 5 correct subtractions: **3 pts**, 2 or 3 correct: **2 pts**, 1 correct: **1 pt**, 0 correct: **0 pt** /3

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 /5

	Has to recall words WITH NO CUE	FACE [ ]	VELVET [ ]	CHURCH [ ]	DAISY [ ]	RED [ ]	Points for UNCUED recall only
Optional	Category cue						
	Multiple choice cue						

ORIENTATION [ ] Date [ ] Month [ ] Year [ ] Day [ ] Place [ ] City /6

© Z.Nasreddine MD www.mocatest.org Normal ≥ 26 / 30

TOTAL /30
Add 1 point if ≤ 12 yr edu

Administered by: _____

Appendix 3

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 (μmol/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/1.73m2)
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 (µmol/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 4

List of Variables after Data Pre-processing

No	Column Name	Blood-based Parameter Name
1	A1_1	Vitamin B12 (pmol/L)
2	A2_1	Serum Homocysteine (µmol/L)
3	A3_1	25-hydroxy Vitamin D (nmol/L)
4	B1_a	Haemoglobin (g/L)
5	B1_a1	RBC (/L)
6	B1_a2	PCV (L/L)
7	B1_a3	MCV (fL)
8	B1_a4	MCH (pg)
9	B1_a5	MCHC (g/L)
10	B1_a6	RDW (%)
11	B1_b	White Cell Count (/L)
12	B1_b1	Neutrophils (/L)
13	B1_b2	Lymphocytes (/L)
14	B1_b3	Monocytes (/L)
15	B1_c	Platelets (/L)
16	B1_d	Glucose (mmol/L)
17	B2_a1	Total Cholesterol (mmol/L)
18	B2_a2	Triglyceride (mmol/L)
19	B2_a3	HDL Cholesterol (mmol/L)
20	B2_a4	LDL Cholesterol (mmol/L)
21	B2_a5	Total Cholesterol/HDL Ratio
22	B2_b1	Sodium (mmol/L)
23	B2_b2	Potassium (mmol/L)
24	B2_b3	Chloride (mmol/L)
25	B2_c1	Urea (mmol/L)
26	B2_c2	Creatinine (umol/L)
27	B2_c4	Uric Acid (mmol/L)
28	B2_c5	Calcium (mmol/L)
29	B2_c6	Corrected Calcium (mmol/L)
30	B2_c7	Phosphate (mmol/L)
31	B2_d1	Total Protein (g/L)
32	B2_d2	Albumin (g/L)
33	B2_d3	Globulin (g/L)
34	B2_d4	Albumin/Globulin ratio
35	B2_d5	Alkaline Phosphatase (U/L)
36	B2_d6	Total Bilirubin (µmol/L)
37	B2_d7	GGT
38	B2_d8	AST
39	B2_d9	ALT
40	B3	C-Reactive Protein

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

Appendix 5

Holdout Method Confusion Matrices

```
Performance Metrics for Logistic Regression:

0.38177

[[108  19  20   0   0   0]
 [ 45  32  12   0   0   0]
 [ 47  18  15   0   0   0]
 [ 26  14  11   0   0   0]
 [ 12  19   3   0   0   0]
 [  3   2   0   0   0   0]]

              precision    recall  f1-score   support

           0       0.45      0.73      0.56       147
           1       0.31      0.36      0.33        89
           2       0.25      0.19      0.21        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      0.21      0.18       406
weighted avg       0.28      0.38      0.32       406
```

```
Performance Metrics for LDA:

0.35714

[[92 25 26  3  1  0]
 [32 32 17  5  2  1]
 [39 20 17  1  1  2]
 [21 15 10  3  1  1]
 [ 8 17  5  3  1  0]
 [ 3  2  0  0  0  0]]

              precision    recall  f1-score   support

           0       0.47      0.63      0.54       147
           1       0.29      0.36      0.32        89
           2       0.23      0.21      0.22        80
           3       0.20      0.06      0.09        51
           4       0.17      0.03      0.05        34
           5       0.00      0.00      0.00         5

    accuracy                           0.36       406
   macro avg       0.23      0.21      0.20       406
weighted avg       0.32      0.36      0.32       406


  Performance Metrics for KNN:

  0.33498

  [[102  20  21   4   0   0]
   [ 47  25  15   1   1   0]
   [ 47  22   9   2   0   0]
   [ 35  12   4   0   0   0]
   [ 17   9   4   4   0   0]
   [  3   1   0   1   0   0]]

                precision    recall  f1-score   support

             0       0.41      0.69      0.51       147
             1       0.28      0.28      0.28        89
             2       0.17      0.11      0.14        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                           0.33       406
     macro avg       0.14      0.18      0.15       406
  weighted avg       0.24      0.33      0.27       406
```

```
Performance Metrics for CART:

0.24138

[[54 27 40 18  7  1]
 [31 21 21  6  8  2]
 [30 21 15 11  3  0]
 [20 14 10  6  1  0]
 [12 11  8  2  1  0]
 [ 0  0  2  2  0  1]]

              precision    recall  f1-score   support

           0       0.37      0.37      0.37       147
           1       0.22      0.24      0.23        89
           2       0.16      0.19      0.17        80
           3       0.13      0.12      0.12        51
           4       0.05      0.03      0.04        34
           5       0.25      0.20      0.22         5

    accuracy                           0.24       406
   macro avg       0.20      0.19      0.19       406
weighted avg       0.24      0.24      0.24       406


Performance Metrics for GNB:

0.38177

[[108  19  20   0   0   0]
 [ 45  32  12   0   0   0]
 [ 47  18  15   0   0   0]
 [ 26  14  11   0   0   0]
 [ 12  19   3   0   0   0]
 [  3   2   0   0   0   0]]

              precision    recall  f1-score   support

           0       0.45      0.73      0.56       147
           1       0.31      0.36      0.33        89
           2       0.25      0.19      0.21        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      0.21      0.18       406
weighted avg       0.28      0.38      0.32       406
```

```
Performance Metrics for SVM:

0.36207

[[147   0   0   0   0   0]
 [ 89   0   0   0   0   0]
 [ 80   0   0   0   0   0]
 [ 51   0   0   0   0   0]
 [ 34   0   0   0   0   0]
 [  5   0   0   0   0   0]]

              precision    recall  f1-score   support

           0       0.36      1.00      0.53       147
           1       0.00      0.00      0.00        89
           2       0.00      0.00      0.00        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                           0.36       406
   macro avg       0.06      0.17      0.09       406
weighted avg       0.13      0.36      0.19       406


Performance Metrics for RFC:

0.38424

[[111  20  16   0   0   0]
 [ 44  36   7   1   1   0]
 [ 50  19   9   0   2   0]
 [ 30   9  12   0   0   0]
 [ 13  17   3   1   0   0]
 [  3   1   0   1   0   0]]

              precision    recall  f1-score   support

           0       0.44      0.76      0.56       147
           1       0.35      0.40      0.38        89
           2       0.19      0.11      0.14        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       0.28      0.38      0.31       406
```

Appendix 6

Cross Validation Confusion Matrices

Logistic Regression:

	precision	recall	f1-score	support
0	0.42	0.90	0.57	68
1	0.41	0.37	0.39	46
2	0.44	0.16	0.23	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	0.21	0.24	0.20	203
weighted avg	0.33	0.42	0.33	203

	precision	recall	f1-score	support
0	0.40	0.69	0.51	68
1	0.34	0.43	0.38	47
2	0.17	0.09	0.12	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	0.23	0.21	0.18	203
weighted avg	0.31	0.35	0.29	203

	precision	recall	f1-score	support
0	0.43	0.77	0.55	69
1	0.37	0.40	0.39	47
2	0.21	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			0.38	203
macro avg	0.20	0.22	0.19	203
weighted avg	0.30	0.38	0.32	203

	precision	recall	f1-score	support
0	0.42	0.86	0.56	69
1	0.33	0.32	0.32	47
2	0.20	0.07	0.10	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			0.38	203
macro avg	0.16	0.21	0.16	203
weighted avg	0.26	0.38	0.29	203

	precision	recall	f1-score	support
0	0.36	0.72	0.49	69
1	0.20	0.22	0.21	46
2	0.15	0.04	0.07	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			0.31	203
macro avg	0.18	0.17	0.14	203
weighted avg	0.25	0.31	0.24	203

LDA:

	precision	recall	f1-score	support
0	0.43	0.79	0.56	68
1	0.46	0.46	0.46	46
2	0.32	0.16	0.21	45
3	0.20	0.04	0.06	27
4	0.00	0.00	0.00	15
5	0.00	0.00	0.00	2
accuracy			0.41	203
macro avg	0.23	0.24	0.21	203
weighted avg	0.35	0.41	0.35	203

	precision	recall	f1-score	support
0	0.41	0.60	0.49	68
1	0.42	0.45	0.43	47
2	0.12	0.09	0.10	45
3	0.11	0.04	0.06	26
4	0.00	0.00	0.00	15
5	0.00	0.00	0.00	2
accuracy			0.33	203
macro avg	0.18	0.20	0.18	203
weighted avg	0.28	0.33	0.30	203

	precision	recall	f1-score	support
0	0.45	0.67	0.53	69
1	0.29	0.34	0.31	47
2	0.36	0.20	0.26	44
3	0.27	0.12	0.16	26
4	0.00	0.00	0.00	16
5	0.00	0.00	0.00	1
accuracy			0.36	203
macro avg	0.23	0.22	0.21	203
weighted avg	0.33	0.36	0.33	203

	precision	recall	f1-score	support
0	0.42	0.70	0.52	69
1	0.36	0.40	0.38	47
2	0.24	0.11	0.15	44
3	0.14	0.04	0.06	27
4	0.00	0.00	0.00	15
5	0.00	0.00	0.00	1
accuracy			0.36	203
macro avg	0.19	0.21	0.19	203
weighted avg	0.30	0.36	0.31	203

	precision	recall	f1-score	support
0	0.36	0.62	0.46	69
1	0.28	0.28	0.28	46
2	0.30	0.13	0.18	45
3	0.30	0.11	0.16	27
4	0.14	0.07	0.09	15
5	0.00	0.00	0.00	1
accuracy			0.33	203
macro avg	0.23	0.20	0.20	203
weighted avg	0.30	0.33	0.29	203

KNN:

	precision	recall	f1-score	support
0	0.36	0.71	0.48	68
1	0.39	0.30	0.34	46
2	0.20	0.09	0.12	45
3	0.07	0.04	0.05	27
4	0.00	0.00	0.00	15
5	0.00	0.00	0.00	2
accuracy			0.33	203
macro avg	0.17	0.19	0.17	203
weighted avg	0.26	0.33	0.27	203

	precision	recall	f1-score	support
0	0.36	0.59	0.44	68
1	0.18	0.17	0.18	47
2	0.21	0.16	0.18	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	0.13	0.15	0.13	203
weighted avg	0.21	0.27	0.23	203

	precision	recall	f1-score	support
0	0.36	0.59	0.45	69
1	0.24	0.21	0.23	47
2	0.18	0.14	0.16	44
3	0.09	0.04	0.05	26
4	0.67	0.12	0.21	16
5	0.00	0.00	0.00	1
accuracy			0.30	203
macro avg	0.26	0.18	0.18	203
weighted avg	0.28	0.30	0.26	203

	precision	recall	f1-score	support
0	0.36	0.68	0.47	69
1	0.31	0.26	0.28	47
2	0.16	0.09	0.12	44
3	0.12	0.04	0.06	27
4	0.00	0.00	0.00	15
5	0.00	0.00	0.00	1
accuracy			0.32	203
macro avg	0.16	0.18	0.15	203
weighted avg	0.25	0.32	0.26	203

	precision	recall	f1-score	support
0	0.37	0.64	0.47	69
1	0.30	0.28	0.29	46
2	0.30	0.13	0.18	45
3	0.17	0.11	0.13	27
4	0.00	0.00	0.00	15
5	0.00	0.00	0.00	1
accuracy			0.33	203
macro avg	0.19	0.19	0.18	203
weighted avg	0.28	0.33	0.28	203

CART:

	precision	recall	f1-score	support
0	0.46	0.53	0.49	68
1	0.36	0.35	0.35	46
2	0.19	0.13	0.16	45
3	0.14	0.19	0.16	27
4	0.18	0.13	0.15	15
5	0.00	0.00	0.00	2
accuracy			0.32	203
macro avg	0.22	0.22	0.22	203
weighted avg	0.31	0.32	0.31	203

	precision	recall	f1-score	support
0	0.44	0.34	0.38	68
1	0.30	0.30	0.30	47
2	0.31	0.27	0.29	45
3	0.08	0.12	0.10	26
4	0.11	0.20	0.14	15
5	0.00	0.00	0.00	2
accuracy			0.27	203
macro avg	0.21	0.20	0.20	203
weighted avg	0.31	0.27	0.28	203

	precision	recall	f1-score	support
0	0.29	0.29	0.29	69
1	0.22	0.23	0.22	47
2	0.19	0.20	0.20	44
3	0.12	0.12	0.12	26
4	0.12	0.06	0.08	16
5	0.00	0.00	0.00	1
accuracy			0.22	203
macro avg	0.16	0.15	0.15	203
weighted avg	0.22	0.22	0.22	203

	precision	recall	f1-score	support
0	0.39	0.36	0.38	69
1	0.26	0.26	0.26	47
2	0.32	0.32	0.32	44
3	0.15	0.15	0.15	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	0.42	0.42	0.42	69
1	0.31	0.22	0.26	46
2	0.24	0.29	0.26	45
3	0.22	0.22	0.22	27
4	0.11	0.13	0.12	15
5	0.00	0.00	0.00	1
accuracy			0.30	203
macro avg	0.22	0.21	0.21	203
weighted avg	0.30	0.30	0.30	203

GNB:

	precision	recall	f1-score	support
0	0.41	0.66	0.51	68
1	0.41	0.24	0.30	46
2	0.33	0.16	0.21	45
3	0.24	0.30	0.26	27
4	0.20	0.13	0.16	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	0.37	0.50	0.42	68
1	0.32	0.21	0.26	47
2	0.11	0.04	0.06	45
3	0.18	0.27	0.22	26
4	0.19	0.27	0.22	15
5	0.00	0.00	0.00	2
accuracy			0.28	203
macro avg	0.20	0.22	0.20	203
weighted avg	0.26	0.28	0.26	203

	precision	recall	f1-score	support
0	0.42	0.64	0.51	69
1	0.16	0.15	0.16	47
2	0.12	0.02	0.04	44
3	0.20	0.27	0.23	26
4	0.08	0.06	0.07	16
5	0.00	0.00	0.00	1
accuracy			0.30	203
macro avg	0.17	0.19	0.17	203
weighted avg	0.24	0.30	0.25	203

	precision	recall	f1-score	support
0	0.43	0.71	0.54	69
1	0.33	0.28	0.30	47
2	0.43	0.14	0.21	44
3	0.16	0.15	0.15	27
4	0.20	0.13	0.16	15
5	0.00	0.00	0.00	1
accuracy			0.36	203
macro avg	0.26	0.23	0.23	203
weighted avg	0.35	0.36	0.33	203

	precision	recall	f1-score	support
0	0.42	0.75	0.54	69
1	0.19	0.09	0.12	46
2	0.38	0.11	0.17	45
3	0.12	0.11	0.12	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

SVM:

	precision	recall	f1-score	support
0	0.33	1.00	0.50	68
1	0.00	0.00	0.00	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			0.33	203
macro avg	0.06	0.17	0.08	203
weighted avg	0.11	0.33	0.17	203

	precision	recall	f1-score	support
0	0.33	1.00	0.50	68
1	0.00	0.00	0.00	47
2	0.00	0.00	0.00	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			0.33	203
macro avg	0.06	0.17	0.08	203
weighted avg	0.11	0.33	0.17	203

	precision	recall	f1-score	support
0	0.34	1.00	0.51	69
1	0.00	0.00	0.00	47
2	0.00	0.00	0.00	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			0.34	203
macro avg	0.06	0.17	0.08	203
weighted avg	0.12	0.34	0.17	203

	precision	recall	f1-score	support
0	0.34	1.00	0.51	69
1	0.00	0.00	0.00	47
2	0.00	0.00	0.00	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			0.34	203
macro avg	0.06	0.17	0.08	203
weighted avg	0.12	0.34	0.17	203

	precision	recall	f1-score	support
0	0.34	1.00	0.51	69
1	0.00	0.00	0.00	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			0.34	203
macro avg	0.06	0.17	0.08	203
weighted avg	0.12	0.34	0.17	203

RFC:

	precision	recall	f1-score	support
0	0.39	0.82	0.53	68
1	0.41	0.33	0.36	46
2	0.22	0.11	0.15	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			0.37	203
macro avg	0.17	0.21	0.17	203
weighted avg	0.27	0.37	0.29	203

	precision	recall	f1-score	support
0	0.43	0.78	0.55	68
1	0.30	0.40	0.35	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	0.38	0.68	0.49	69
1	0.23	0.28	0.25	47
2	0.18	0.09	0.12	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			0.32	203
macro avg	0.13	0.17	0.14	203
weighted avg	0.22	0.32	0.25	203

	precision	recall	f1-score	support
0	0.43	0.70	0.53	69
1	0.36	0.49	0.41	47
2	0.19	0.11	0.14	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	0.33	0.22	0.19	203
weighted avg	0.40	0.38	0.32	203

	precision	recall	f1-score	support
0	0.36	0.71	0.47	69
1	0.22	0.24	0.23	46
2	0.00	0.00	0.00	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			0.30	203
macro avg	0.26	0.17	0.14	203
weighted avg	0.25	0.30	0.22	203