PREDICTING MORTALITY OF MALAYSIAN PATIENTS WITH ACUTE CORONARY SYNDROME (ACS) SUBTYPES USING MACHINE LEARNING AND DEEP LEARNING APPROACHES

MUHAMMAD FIRDAUS BIN AZIZ

FACULTY OF SCIENCE
UNIVERSITI MALAYA
KUALA LUMPUR

2022

PREDICTING MORTALITY OF MALAYSIAN PATIENTS WITH ACUTE CORONARY SYNDROME (ACS) SUBTYPES USING MACHINE LEARNING AND DEEP LEARNING APPROACHES

MUHAMMAD FIRDAUS BIN AZIZ

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

INSTITUTE OF BIOLOGICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITI MALAYA
KUALA LUMPUR

2022

UNIVERSITI MALAYA
ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **MUHAMMAD FIRDAUS BIN AZIZ**

Matric No: **17051443/4**

Name of Degree: **DOCTOR OF PHILOSOPHY**

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

PREDICTING MORTALITY OF MALAYSIAN PATIENTS WITH ACUTE CORONARY SYNDROME (ACS) SUBTYPES USING MACHINE LEARNING AND DEEP LEARNING APPROACHES

Field of Study:

BIOINFORMATICS

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date: 29/7/2022

Subscribed and solemnly declared before,

Witness's Signature Date: 29 July 2022

Name:

Designation:

PREDICTING MORTALITY OF MALAYSIAN PATIENTS WITH ACUTE CORONARY SYNDROME (ACS) SUBTYPES USING MACHINE LEARNING AND DEEP LEARNING APPROACHES

ABSTRACT

The conventional risk score for predicting short- and long-term mortality following Acute Coronary Syndrome (ACS) is typically not population-specific and does not accommodate for Asian patients. The purpose of this study is to use machine learning (ML) and deep learning (DL) algorithms to predict and identify variables linked to short and long-term mortality in Asian STEMI and NSTEMI/UA patients and to compare these results to a conventional risk score. Model development for STEMI: in-hospital (6299 patients), 30-days (3130 patients), and 1-year (2939 patients) and NSTEMI/UA: in-hospital (4771 patients), 30-days (2402 patients), and 1-year (2304 patients) datasets was done using the National Cardiovascular Disease Database (NCVD) Malaysia registry of a multi-ethnic, heterogeneous Asian ACS population. 50 variables were considered for STEMI and 39 for NSTEMI/UA. ML algorithms were used to examine significant variables utilising feature selection methods. The ML feature selection approach was then used to develop ML and DL models using all and selected variables, which were then compared to the Thrombolysis in Myocardial Infarction (TIMI) score. For STEMI patients, the best ML model, a Support Vector Machine (SVM) classifier with sequential backward elimination (SBE) selected variables, produced AUC values of 0.88 for in-hospital, 0.90 for 30 days, and 0.84 for 1 year, while the best model for NSTEMI/UA patients produced AUC values of 0.85 for in-hospital, 0.87 for 30 days, and 0.80 for 1-year mortality prediction. The same variables were then used to create the best DL model for STEMI (AUC 0.96 in-hospital, 0.93 for 30 days, and 0.90 for 1-year mortality prediction) and NSTEMI/UA (AUC 0.97 in-hospital,

0.91 for 30 days, and 0.88 for 1-year mortality prediction). TIMI risk score reported lower performance for STEMI (In-hospital: AUC=0.81, 30 days: AUC=0.80 and 1-year: AUC=0.76) and NSTEMI/UA patients (In-hospital: AUC=0.42, 30 days: AUC=0.49 and 1-year: AUC=0.42) as compared to ML and DL algorithms. Age, heart rate, Killip class, fasting blood glucose, and diuretics were found to be the common variables across the three time points in the STEMI dataset, whereas age, heart rate, Killip class, and intake of Low-molecular-weight heparin (LMWH) were found to be the common variables in the NSTEMI/UA dataset. When compared to the TIMI risk score, both ML and DL were better at classifying ACS patients in a multi-ethnic population. ML enables the identification of distinct variables in Asian populations to improve mortality prediction. In the future, continuous testing and validation will enable improved risk classification, possibly modifying management and results.

Keyword: STEMI, NSTEMI/UA, population-specific, deep learning, machine learning, mortality prediction, Asian.

MERAMAL KEMATIAN PESAKIT SUBJENIS SINDROM KORONARI AKUT (SKA) DI MALAYSIA MENGGUNAKAN PENDEKATAN PEMBELAJARAN MESIN DAN PEMBELAJARAN DALAM

ABSTRAK

Skor risiko konvensional untuk meramal kematian jangka pendek dan jangka panjang berikutan Sindrom Koronari Akut (ACS) biasanya tidak khusus kepada sesebuah populasi dan tidak sesuai untuk pesakit Asia. Tujuan kajian ini adalah untuk menggunakan algoritma pembelajaran mesin (ML) dan pembelajaran dalam (DL) untuk meramal dan mengenal pasti pembolehubah yang dikaitkan dengan kematian jangka pendek dan jangka panjang dalam pesakit STEMI dan NSTEMI/UA di Asia, serta untuk membandingkan keputusan ini dengan skor risiko konvensional. Pembangunan model untuk STEMI: dalam hospital (6299 pesakit), 30 hari (3130 pesakit), dan 1 tahun (2939 pesakit) serta NSTEMI/UA: dalam hospital (4771 pesakit), 30 hari (2402 pesakit), dan 1 tahun (2304 pesakit) telah dilaksanakan menggunakan Pangkalan Data National Cardiovascular Disease (NCVD) Malaysia bagi populasi ACS Asia yang berbilang etnik dan heterogen. 50 pembolehubah telah dipertimbangkan untuk set data STEMI dan 39 untuk set data NSTEMI/UA. Algoritma ML digunakan untuk mengkaji pembolehubah penting menggunakan kaedah pemilihan ciri. Pendekatan pemilihan ciri ML kemudiannya digunakan untuk membina model ML dan DL menggunakan semua pembolehubah terpilih, yang kemudiannya dibandingkan dengan skor Thrombolysis in Myocardial Infarction (TIMI). Bagi pesakit STEMI, model ML terbaik, Mesin Vektor Sokongan (SVM) dengan pembolehubah terpilih menggunakan penghapusan berurutan dari belakang (SBE), menghasilkan nilai AUC sebanyak 0.88 untuk dalam hospital, 0.90 untuk 30 hari dan 0.84 untuk 1 tahun, manakala model terbaik untuk pesakit NSTEMI/UA menghasilkan nilai AUC sebanyak 0.85 untuk

dalam hospital, 0.87 untuk 30 hari dan 0.80 untuk ramalan kematian 1 tahun. Pembolehubah yang sama kemudiannya digunakan untuk membina model terbaik DL untuk set data STEMI (AUC 0.96 dalam hospital, 0.93 untuk 30 hari dan 0.90 untuk ramalan kematian 1 tahun) dan NSTEMI/UA (AUC 0.97 dalam hospital, 0.91 untuk 30 hari, dan 0.88 untuk ramalan kematian 1 tahun). Skor risiko TIMI melaporkan prestasi yang lebih rendah untuk STEMI (Dalam hospital: AUC=0.81, 30 hari: AUC=0.80 dan 1 tahun: AUC=0.76) dan pesakit NSTEMI/UA (Dalam hospital: AUC=0.42, 30 hari: AUC =0.49 dan 1- tahun: AUC=0.42) berbanding algoritma ML dan DL. Umur, kadar denyutan jantung, kelas Killip, glukosa darah puasa dan pengambilan ubat diuretik telahpun dipilih sebagai pembolehubah yang serupa merentas tiga titik masa ramalan kematian dalam set data STEMI, manakala umur, kadar denyutan jantung, kelas Killip dan pengambilan ubat "low-molecular weight heparin" (LMWH) didapati sebagai pembolehubah yang sama dalam dataset NSTEMI/UA bagi tiga titik masa ramalan kematian. Jika dibandingkan dengan skor risiko TIMI, kedua-dua ML dan DL melaksanakan tugas yang lebih baik dalam mengklasifikasikan pesakit ACS dalam populasi berbilang etnik. ML membolehkan pengenalpastian pembolehubah yang berbeza dalam populasi Asia untuk meningkatkan kebolehan ramalan kematian. Pada masa hadapan, ujian dan pengesahan berterusan akan membolehkan klasifikasi risiko yang lebih baik terhadap algoritma yang tersedia ada ini, dan mungkin juga dapat mengubah suai pengurusan dan keputusan pesakit ACS.

Kata kunci: STEMI, NSTEMI/UA, khusus populasi, "machine learning", "deep learing", ramalan kematian, Asia.

ACKOWLEDGEMENTS

I would like to express my sincere gratitude and appreciation to both my supervisors, Dr. Sorayya Malek and Prof. Sazzli Kasim for giving me the opportunity to pursue my doctoral studies. I am deeply grateful for your guidance, patience, support, and constant encouragement throughout my study and the preparation of this thesis. Both of you have provided me with a conducive working environment and are always accessible whenever I need help. The knowledge that both of you have shared along this journey will forever be in my debt and only God could repay both of you with endless blessings and fortunes.

I would also like to thank the National Heart Association Malaysia (NHAM) for providing the data for my study from the National Cardiovascular Disease Database (NCVD). I am forever grateful for the help I received from the association in preparing the data for my study and to be given the opportunity to work with the national data.

To my family members, words can never describe all the sacrifices that all of you have done for me over the years in terms of finances and time. As I worked toward earning my doctoral degree, I am grateful to both my parents for their unwavering belief in my ability to achieve all of my goals and for their constant support in keeping me mentally well throughout the process. Thank you for letting me do the thing I love the most in the world, which is to learn. I doubt I will ever be able to convey my appreciation fully to any of you.

Last but not least, I would also like to thank my friends who have supported me since day one of this journey with endless words of encouragement, the time spent, and the abundance of ways to make me feel like a human again which always put me back on track on why I started this journey at the first place. I wish nothing but gratefulness in each one of you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACE	:	Angiotensin-Converting Enzyme
acei	:	ACE inhibitor
ACS	:	Acute Coronary Syndrome
AI	:	Artificial Intelligence
AIC	:	Akaike Information Criterion
ANN	:	Artificial Neural Network
antiarr	:	Anti-arrhythmic agent
ARB	:	Angiotensin II Receptor Blocker
arb	:	Angiotensin II receptor blocker
asa	:	Aspirin
AUC	:	Area Under the ROC Curve
bb	:	Beta blockers
BMI	:	Body Mass Index
bpdias	:	Diastolic blood pressure
bpsys	:	Systolic blood pressure
CABG	:	Coronary Artery Bypass Grafting
CAD	:	Coronary Artery Disease

calcantagonist	:	Calcium antagonists
canginapast2wk	:	Chronic angina from the past two weeks
ccap	:	Documented CAD
ccerebrovascular	:	Cerebrovascular disease
cdm	:	History of diabetes
cdys	:	History of dyslipidaemia
cheartfail	:	History of heart failure
chpt	:	History of hypertension
clung	:	Chronic lung disease
cmi	:	History of MI
cpremcvd	:	Family history of premature cardiovascular disease
crenal	:	Chronic renal disease
CVD	:	Cardiovascular Disease
diuretic	:	Diuretics
DL	:	Deep Learning
DT	:	Decision Tree
ECG	:	Electrocardiogram
ecgabnormlocational	:	Anterior leads: V1 to V4
ecgabnormlocationil	:	Inferior leads: II, III, aVF

ecgabnormlocationll	:	Lateral leads: I, aVL, V5 to V6
ecgabnormlocationrv	:	Right ventricle: ST elevation in lead V4R
ecgabnormlocationtp	:	True posterior: V1, V2
ecgabnormstylestdep	:	ST segment depression ≥0.5mm in ≥2 contiguous leads
ecgabnormstylestelev1	:	ST segment elevation ≥1mm in ≥2 contiguous limb leads
ecgabnormstylestelev2	:	ST segment elevation ≥2mm in ≥2 contiguous frontal leads or chest leads
ecgabnormtypebbb	:	Bundle branch block
ecgabnormtypetwave	:	T-wave inversion ≥1mm
ED	:	Emergency department
EHR	:	Electronic health record
EN	:	Elastic Net
fbg	:	Fasting blood glucose
fbstatus	:	Fibrinolytics status
FN	:	False negative
FP	:	False positive
gpri	:	Glycoprotein receptor inhibitor
HDL	:	High-Density Lipoprotein
heartrate	:	Heart rate

heparin	:	Unfractionated heparin
history	:	History of MI and CAD
ICU	:	Intensive Care Unit
IHD	:	Ischaemic heart diease
insulin	:	Insulin
killipclass	:	Killip class
kNN	:	k-Nearest Neighbour
LDL	:	Low-Density Lipoprotein
lipidla	:	Other lipid-lowering agents
lmwh	:	Low Molecular Weight Heparin
LR	:	Logistic Regression
MAR	:	Missing at random
MCAR	:	Missing completely at random
MI	:	Myocardial infarction
ML	:	Machine Learning
MLP	:	Multilayer Perceptron
MNAR	:	Missing not at random
NB	:	Naïve Base
NCVD	:	National Cardiovascular Disease Database

NHAM	:	National Heart Association Malaysia
NHMS	:	National Heart Morbidity Survey
NPV	:	negative predictive value
NSTEMI	:	Non ST-Elevation Myocardial Infarction
oralhypogly	:	Oral hypoglycaemic agent
PCI	:	percutaneous coronary intervention
pci	:	Percutaneous Coronary Intervention
PMM	:	Predictive mean matching
PPV	:	Positive predictive value
ptagenotification	:	Age
ptrace	:	Race
ptsex	:	Gender
RBF	:	Radial basis function
ReLU	:	Rectified Linear Unit
RF	:	Random Forest
RFE	:	Recursive Feature Elimination
ROSE	:	Random Over Sampling Examples
SBE	:	Sequential Backward Elimination
smokingstatus	:	Smoking status

SOM	:	Self-Organising Map
STEMI	:	ST-Elevation Myocardial Infarction
SVM	:	Support Vector Machine
tc	:	Total cholesterol
tg	:	Triglycerides
TN	:	True positive
TP	:	True negative
UA	:	Unstable Angina
WHO	:	World Health Organization

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 Background of the study

The title of the study is 'Predicting Mortality of Malaysian Patients with Acute Coronary Syndrome (ACS) Subtypes using Machine Learning and Deep Learning Approaches'. As the title suggests, this study introduces new methods of predicting the risk of ACS subtypes patient mortality; STEMI and NSTEMI/UA using machine learning (ML) algorithms in the short and long-term period using a multi-ethnic Malaysian registry.

Cardiovascular disease (CVD) is the primary cause of death in the world. This is despite recent advances in health care systems, pharmacotherapy including fibrinolytic and revascularization options such as percutaneous coronary intervention (PCI) with the use of drug-eluting stents and Coronary Artery Bypass Surgery (CABG) (Reddy et al., 2015; WHO, 2020). Acute coronary syndrome (ACS) is a type of CVD that is defined as unstable angina (UA), ST-Elevation Myocardial Infarction (STEMI), and non-ST-Elevation Myocardial Infarction (NSTEMI) (Damman et al., 2017). According to World Health Organization (WHO), cardiovascular disease is responsible for 16% of the world's mortality and this disease has been the highest contributor to the increase in death since the year 2000, rising by more than 2 million to 8.9 million deaths in 2019 (WHO, 2020). CVD was also the leading cause of death in Asia in 2019, causing 10.8 million deaths, which were approximately 35% of the total deaths in Asia and 39% in South-East Asia specifically (IHME, 2020). Ischemic heart disease remained the leading cause of death in Malaysia in 2019, accounting for 15% of the 109,164 medically certified deaths (DOSM, 2020). 20-25% of all deaths in public hospitals are attributed to coronary artery diseases (CAD) with a higher mortality rate reported for 30-day mortality following ACS (Hoo et al., 2016).

It is critical to recognise a patient with ACS as soon as possible since the right treatment and monitoring can significantly improve the patient's prognosis. Identifying a patient's risk of complications or death is necessary to make decisions on escalating or de-escalation of treatment. To do this, risk scores derived from cohorts of patients are used in creating models. Risk scores such as Thrombolysis in Myocardial Infarction (TIMI) or the Global Registry of Acute Cardiac Events (GRACE) are frequently used to predict the mortality risks associated with ACS. These traditional risk scores are derived from research involving mostly Caucasian individuals and just a small number of Asian patients (Peng et al., 2017). Younger patients with myocardial infarction (MI), a larger burden of diabetes mellitus, hypertension, and renal failure, as well as a higher prevalence of delayed presentation for medical care, are more common in Asian nations. Premature mortality in Asia (39%) significantly greater than premature CVD deaths in the United States (23%), Europe (22%), and the rest of the world (34%) (IHME, 2020). Malaysia is unique in that it is heterogeneous owing to innate genetic differences in a multi-ethnic population that is already diversified. Any study of disease among distinct ethnic groups is challenged by other factors such as lifestyle, geography, and socioeconomic level. Ethnicity variations will result in different disease rates and risk variables from one country to another (Anand, 1999; Peng et al., 2017). This demonstrates that these risk scores are not population-specific and may be unable to account for regional differences in disease burden, healthcare resources, and management options. The inability to accurately classify patients into appropriate risk scores may underline the differences in outcomes. These factors may explain the discrepancy of outcomes amongst Asian patients being treated for ACS.

When TIMI scores were examined between Asians and Caucasians, it was shown that Asians had a greater incidence of STEMI with similar mortality risk. This disparity is difficult to explain, especially given the fact that Asian patients have a greater disease

burden (Selvarajah et al., 2012). Aside from that, traditional cardiovascular disease risk assessment models presume a linear relationship between risk factors and clinical outcomes, resulting in an oversimplification of an actually complicated relationship. Models that take into account these various risk variables and outcomes, such as the application of Artificial Intelligence (AI), are needed (Kim & Groeneveld, 2017; Obermeyer & Emanuel, 2016).

AI techniques such as ML and Deep Learning (DL) may play a key role in the advancement of cardiovascular therapy by facilitating precision cardiovascular investigations (Krittanawong et al., 2017). Traditional regression-based prediction models of CVD occurrences are used to create conventional risk scores, while AI techniques are known as strategies to overcome their shortcomings. The key aims of AI-based mortality prediction models are to establish connections between diverse illnesses, achieve high prediction accuracy, and interpret missing and outlier data well. It is also feasible to undertake data analysis with dependent variables on limited and incomplete training data sets, which is a drawback of the regression-based model (logistic regression and Cox proportional hazard regression models) (Hsieh et al., 2019).

ML algorithms are equipped to handle complex data and provided accurate risk-prediction models at the population-specific level (Wallert et al., 2017). ML algorithms also requires feature selection methods to achieve higher model performance accuracy (Chen & Ishwaran, 2012). To conduct ML, there are three things needed, namely, input data points, examples of the expected output, and performance metrics (Chollet, 2018).

On the other hand, DL enables computational models that consisted of multiple layers of processing to learn data representations with multiple abstraction levels. According to LeCun et al. (2015), DL does not perform feature selection but instead uses feature learning. Feature learning can only learn all of the variables that have been presented and

do the tasks that have been assigned to it, such as classification and detection, in order to gain relevant variables that can be used to predict the outcome. However, finding a limited number of variables related to mortality is critical for identifying characteristics of high-risk patients in clinical practice and improving patient care. Hence, this study employed the features selected by the best ML model into the DL model for the precision of the outcome.

There have been some previous studies that employed both ML and DL algorithms to develop predictive models in predicting the risk mortality of patients with ACS from countries such as United States of America (Frizzell et al., 2017), Sweden (Wallert et al., 2017), Israel (Shouval et al., 2017), Korea (Lee et al., 2021; Sherazi et al., 2021) and China (Li et al., 2020). The results of their studies showed that DL or ML models that they developed outperformed the conventional risk scores such as TIMI and GRACE risk scores by achieving higher AUC values. Some of the common algorithms used for the ACS patient's mortality prediction are Naïve Bayesian Network, Random Forest, Support Vector Machine, Gradient Boosting, Extreme Gradient boosting, Ada Boost, Alternating Decision Tree, and Deep Neural Network.

Based on the promising performance of the ML and DL algorithms in previous studies, these algorithms were employed as an alternative to the present conventional risk scoring approach in the real world. This study will aid in the management of patients as well as potentially improve patient outcomes and reduce treatment costs for both patients and hospitals.

Advances in the identification and management of those high-risk patients with CVD risk factors have resulted in substantial reductions in mortality rates in high-income countries (Leong et al., 2017). This finding emphasises the importance of developing population strategies that focus on reducing the incidence of death, as well as the primary and secondary prevention of CVD.

Hence, obvious economic constraints must be acknowledged, and the capital expense of these high-priority approaches must be viewed in a broader context. Compatibility studies can help pick which treatment and prevention activities to adopt given the restricted financial resources for healthcare. This approach will help make healthcare services more efficient in the future.

In evaluating the effectiveness of the medical intervention, several outcome points may be considered. Surrogate outcomes such as patients' biomarker readings can reflect a short-term change in response to treatment. This however may not matter if it fails to improve health care or individual finances. As such, death is useful as a hard outcome. It is well documented by Malaysian National Registration Department, captured in the Cardiovascular Registry, and commonly used as an outcome measure to benchmark standards of care. Using in-hospital death reflects the care received by patients by the healthcare system prior to arriving at the health care facility (pre-hospital care). This reflects the strength of community care, emergency medical services as well as acute medical care (emergency departments). The use of 30 days death reflects care received during the acute phase in the hospital and reflects the quality of the hospital, sufficient manpower as well as access to interventions that could save lives. 1-year death reflects the care received by patients during the post-acute care in the community. It is during this period that the intervention directed to medication adherence and lifestyle intervention are reflected.

Many middle- and low-income countries will have significant challenges in applying proven treatments. Cost and lack of personnel negate the implementation of optimised strategies to improve health (Joseph et al., 2017). These challenges affect health systems, practitioners, or patients, resulting in large evidence-practice gaps. Thus, health systems must integrate research findings into practice in order to reduce the burden of CVD. To

achieve maximum potential, the best available evidence must be combined with an efficient governance framework to close the knowledge gap, and this study may aid in addressing such issues and improving the country's healthcare system administration.

1.2 Overview of the study

This study was carried out since there has been limited research on ML-based mortality prediction models in clinical ACS patients, particularly in a multi-ethnic country such as Malaysia, implying the need to assess and forecast mortality. Numerous conventional risk scores are still used to objectively assess a patient's risk-benefit ratio following ACS occurrence, but the development of the risk stratification method is not population-specific since the majority of risk scores were developed nearly two decades ago and were based on a predominantly Caucasian population, and as different cultures and lifestyles may influence the predictors and outcome of a patient from another region of the world. Hence, with the growing use of ML methods as a classifier in the medical field, this study considers it as an alternative in developing a risk prediction method for multi-ethnic Malaysian patients with ACS subtypes. The ML-based prediction model additionally adjusts the prediction models to be more population-specific by combining pre-selected variables from the current traditional technique with other relevant predictors.

These past few years, the emergence of a branch of ML method called deep learning (DL) has been widely used in the medical field, but it has a small number of applications in predicting mortality of ACS patients, especially in Asia counterpart. Hence, this study also used DL to compare its effectiveness together with ML methods for predicting mortality of ACS patients in Malaysia.

Patients with ACS subtypes had their mortality predicted at three separate time points: in-hospital, 30 days, and 1-year. In the future, this algorithm will be tested in a real-world setting using a web-based system.

1.3 Problem statement

The best method to minimise ACS-related deaths in Malaysia is to take preventative measures and provide appropriate treatment to ACS patients (Ho et al., 2008). This alternative can be assisted by the building of a predictive model to calculate the mortality of patients while being treated in the hospital.

Extension of the medical field with the aid of the AI method may be an evolutional way of predicting mortality after the ACS. The main purpose of developing the model is to assist medical practitioners in Malaysia in taking further precautions after a patient's first ACS to prevent a second attack that could result in death, based on fitting data of coronary illnesses in the Malaysian population. Hence the problem statement to this study arises from these statements which are:

- Inability to identify the factors that give a high contribution to ACS occurrence.

- Lack of precise prediction model of patient mortality after ACS for Malaysian population in medical practices.

- Limitation in the existing conventional risk score.

1.4 Research questions

RQ1: What are the factors affecting the mortality of ACS patients after their ACS occurrence?

RQ2: Which ML and DL predictive model is suitable to predict mortality in ACS patients?

RQ3: What is the difference in performance between the ML and DL predictive model with the existing conventional predictive model used in Malaysia?

1.5 Research objectives

This study aims to develop a prediction model using ML methods to predict the mortality of the patients after their ACS episode using a list of predictors. The objectives of the study are as follows:

- To identify factors affecting the mortality of ACS patients after their ACS occurrence using ML methods.
- To build a predictive model that is suitable to predict mortality after the first ACS for the Malaysian population.
- To compare the performance of the ML and DL predictive model with the conventional existing model.

1.6 Scope of research

This study focused on the ACS subtypes STEMI and NSTEMI/UA, as these are two categories that account for the majority of CVD mortality. The sample dataset in this study consisted of demographic, patient's status before the event, clinical representation, baseline

investigation, electrocardiography, treatment, and pharmacological therapy data from Malaysians diagnosed with ACS from the National Cardiovascular Disease Database (NCVD) from 2006 to 2016. This is to ensure the features observed, and the predictive model is suitable for the use of the Malaysian population. The outcome of the dataset is the survival or non-survival of patients during the period of in-hospital, 30 days and 1 year from the ACS occurrence. Three ML models were developed using Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR), and two feature selection methods were used for each of the ML models which are Sequential backward elimination (SBE) and Recursive Feature Elimination (RFE). The development of the DL model were based on the variables selected by the best ML model because DL cannot perform feature selection. The results of all of these models were then be compared to the TIMI risk score, a widely used risk scoring system in Malaysia. The best performing models built using ML and DL will be chosen to be incorporated into the proposed system among the developed models.

1.7 Thesis Outline

Chapter 1: Introduction. This chapter describes the cardiovascular event in Malaysia and how crucial it is to overcome the issue. It also expresses the current methods being used and how obscure it is to the Malaysian population. A few alternative methods of prevention are being introduced and considered based on the current technology. Research problems, objectives, and scopes are clearly stated in this chapter.

Chapter 2: Literature Review. This section entails the study's full details. It presents an overview of STEMI and NSTEMI/UA, as well as their conventional risk scores and the risk variables that influence STEMI and NSTEMI/UA patient mortality. It also assesses

previous research and studies on ML and DL model development for heart disease patients, as well as existing mortality prediction systems. A background review of the proposed methodologies is also included in this chapter, which encompasses preprocessing, development, parameter tuning, model performance evaluation, and statistical analysis.

Chapter 3: Methodology. This chapter elaborates on the methodologies and steps to develop the proposed ML and DL models together with the evaluation of the best models' performances.

Chapter 4: Results. This chapter summarised the prediction outcome and the performance of each ML and DL method.

Chapter 5: Discussion. This chapter evaluates the performance of the best model and compares the prediction results with the TIMI risk score. It also discusses the predictors that lead to mortality of STEMI and NSTEMI/UA patients with the predictors from other risk scores. Secondary analyses of the best prediction models are also discussed in this chapter.

Chapter 6: Conclusion. This chapter summarised the whole study, including the strengths, limitations, and further improvements to this study.

CHAPTER 2: LITERATURE REVIEW

2.1 Acute Coronary Syndrome (ACS)

According to the most recent WHO data, 34,766 people died in Malaysia from coronary heart disease in 2018, accounting for 24.69% of all deaths (WHO, 2020). With a death rate of 157.39 per 100,000 inhabitants, Malaysia is ranked 64th in the world. CVD is responsible for 16% of global mortality, and it has been the leading cause of death increase since 2000, with more than 2 million deaths in 2019 compared to 8.9 million in 2000. In 2019, ischemic heart disease (IHD) was the largest cause of mortality in Asia, accounting for 10.8 million deaths, or about 35% of all deaths in Asia and 39% in South-East Asia (IHME, 2020). According to Malaysia's Department of Statistics, IHD was the biggest cause of death in 2019, accounting for 15% of the country's 109,164 medically certified deaths. According to the National Heart Association Malaysia (NHAM)'s Annual Report of the Acute Coronary Syndrome (ACS) Registry, 2014–2015, 46.1% of all patients admitted with ACS in 2014–2015 had ST-elevation myocardial infarction (STEMI), 25.2% had non-STEMI (NSTEMI), and 28.7% had unstable angina (UA) (Ahmad, 2017).

IHD presents itself clinically as ACS. A condition in which a segment of the heart muscle is unable to function properly due to a reduction in blood flow in the coronary arteries, resulting in cell death, is known as ACS (Amsterdam et al., 2014). Fatty deposits in and on the walls of the coronary arteries - the arteries that supply the cell with oxygen and nutrients – cause this to happen (Acharya et al., 2017). Even if the diminished blood flow does not result in myocardial necrosis, it changes the function of the heart and leads to MI.

MI is a subset of the ACS, causes the death of cardiac muscle indicated by elevated cardiac biomarkers levels in the setting of acute ischemia (Barstow et al., 2017). The patients usually presented with chest pain. Various combinations of the chest, upper extremity, jaw, or epigastric discomfort during exertion or at rest are also possible ischemic signs. Acute MI causes discomfort that lasts at least 20 minutes. Discomfort is frequently widespread, not localised, positional, or affected by the movement of the region, and it might be followed by dyspnoea, diaphoresis, nausea, or syncope. These symptoms are not exclusive to myocardial ischaemia and might be mistaken for gastrointestinal, neurological, pulmonary, or musculoskeletal problems. MI might present with atypical symptoms or possibly go unnoticed, with only an ECG, biomarker elevations, or cardiac imaging to detect it (Thygesen et al., 2007).

The diagnosis of MI can only be confirmed based on the presence of myocardial injury (myocardial necrosis) in a clinical setting together with myocardial ischemia indicated by the cardiac biomarker value fluctuation (preferably cardiac troponin) with at least one value above 99th percentile of upper reference limit (Thygesen et al., 2018). In addition, there should be at least one of the following:

- Clinical history is consistent with chest pain or ischaemic origin.
- ECG changes of new ST-T changes or new left bundle branch block (LBBB) indicative of new ischaemia.
- Development of pathological Q waves in the ECG.
- Imaging evidence of new loss of viable myocardium or new regional wall motion abnormality.
- Identification of an intracoronary thrombus by angiography or autopsy.

After a few revisions on MI definitions, medical practitioners have divided MI into five smaller subgroups based on pathological and clinical features (Thygesen, 2007).

Table 2.1:Types of MI.

Type		Description
1: Spontaneous MI		Spontaneous MI is related to atherosclerotic plaque rupture, ulceration, fissuring, erosion, or dissection with resulting intraluminal thrombus in one or more of the coronary arteries leading to decreased myocardial blood flow or distal platelet emboli with ensuing myocyte necrosis.
2: MI secondary to an ischaemic imbalance		In instances of myocardial injury with necrosis where a condition other than CAD contributes to an imbalance between myocardial oxygen supply
3: MI resulting in death when biomarker values are unavailable		Cardiac death with symptoms suggestive of myocardial ischaemic and presumed new ischaemic ECG changes or new LBBB, but death occurring before blood samples could be obtained before cardiac biomarker could rise, or in rare cases, cardiac biomarkers were not collected
4	a: MI related to PCI	MI associated with PCI is arbitrarily defined by the elevation of cardiac troponin (cTn) values > 5 x 99th percentile upper reference limits (URL) in patients with normal baseline values 99th percentile URL) or a rise of cTn values > 20% if the baseline values are elevated and are stable or falling.
	b: MI related to stent thrombosis	MI associated with stent thrombosis is detected by coronary angiography or autopsy in the setting of myocardial ischaemia and with a rise and/or fall of cardiac biomarkers values with at least one value above the 99th percentile URL.
5: MI related to CABG		MI associated with CABG is arbitrarily defined by the elevation of cardiac biomarker values 10 x 99th percentile URL in patients with normal baseline cTn values (99th percentile URL).

2.1.1 Classification of ACS

MI is a subset of ACS which is defined based on its spectrum of clinical presentation upon admission or cardiac arrest. Patients who are diagnosed with MI are further classified into ST-Elevation Myocardial Infarction (STEMI) and non-ST Elevation ACS, which

comprises of unstable angina (UA) and non-ST-Elevation Myocardial Infarction (NSTEMI). These subsets of ACS differ from each other based on the three parameters evaluated prior to diagnosis; presenting signs and symptoms, rapid electrocardiography (ECG), and cardiac biomarkers.

STEMI is defined as cardiac ischemia symptoms characteristic with persistent ST-segment elevation in the resting ECG supported by the presence of raised cardiac biomarker (O'gara et al., 2013). STEMI chest pain starts suddenly and lasts for more than thirty minutes. It commonly starts in the middle of the chest and radiates down the left arm or to the jaw. It might happen when patients are at rest or when doing something active (MOH, 2019). Rapid access to coronary revascularization techniques is advocated and reducing door-to-balloon times for these patients remains a top focus. The chosen cardiac biomarker in individuals with clinical symptoms and an ECG diagnosis of STEMI is CK-MB. Troponins are not needed in this case because there is already evidence of myocardial damage on the ECG.

If there was no ST-elevation in the ECG, the patients are to be diagnosed either with NSTEMI or UA. NSTEMI is the persistent symptoms with elevated cardiac biomarker without ST-segment elevation while UA produces the same suggestive symptoms of cardiac ischemia but no elevation in the cardiac troponin level. In most individuals with NSTEMI/UA, chest pain is a presenting symptom. Retrosternal, central, or left chest pain or discomfort is common, and it might radiate to the mouth or down the upper limb. Crushing, pressing, or burning are examples of natural processes. The intensity of the pain varies (MOH, 2011). Troponins - cTn (I or T) are the recommended biomarkers if the clinical symptoms and ECG are suggestive but not definitive of MI.

This more comprehensive definition of MI, which incorporates updated cardiac biomarkers and imaging tools, improves the sensitivity of MI diagnosis (Salomaa et al., 2005). Figure 2.1 below shows the summary of patient MI classification upon presentation and Figure 2.2 depicts the ECG pattern of the different ACS subtypes.

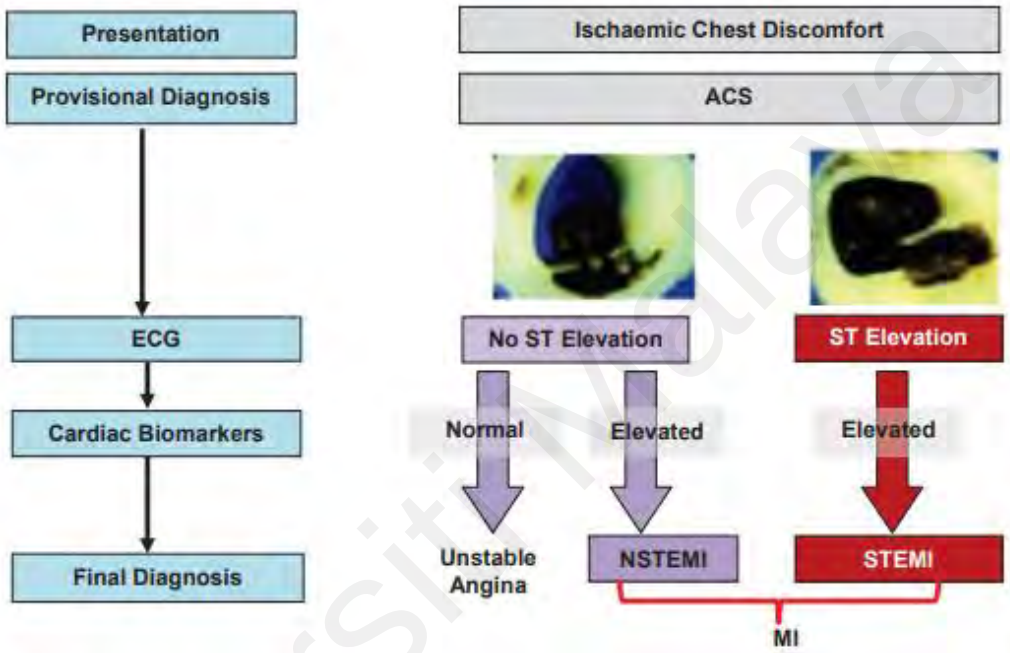


Figure 2.1: Procedure for assessing patients suffering from chest pain. (Antman et al., 2008).

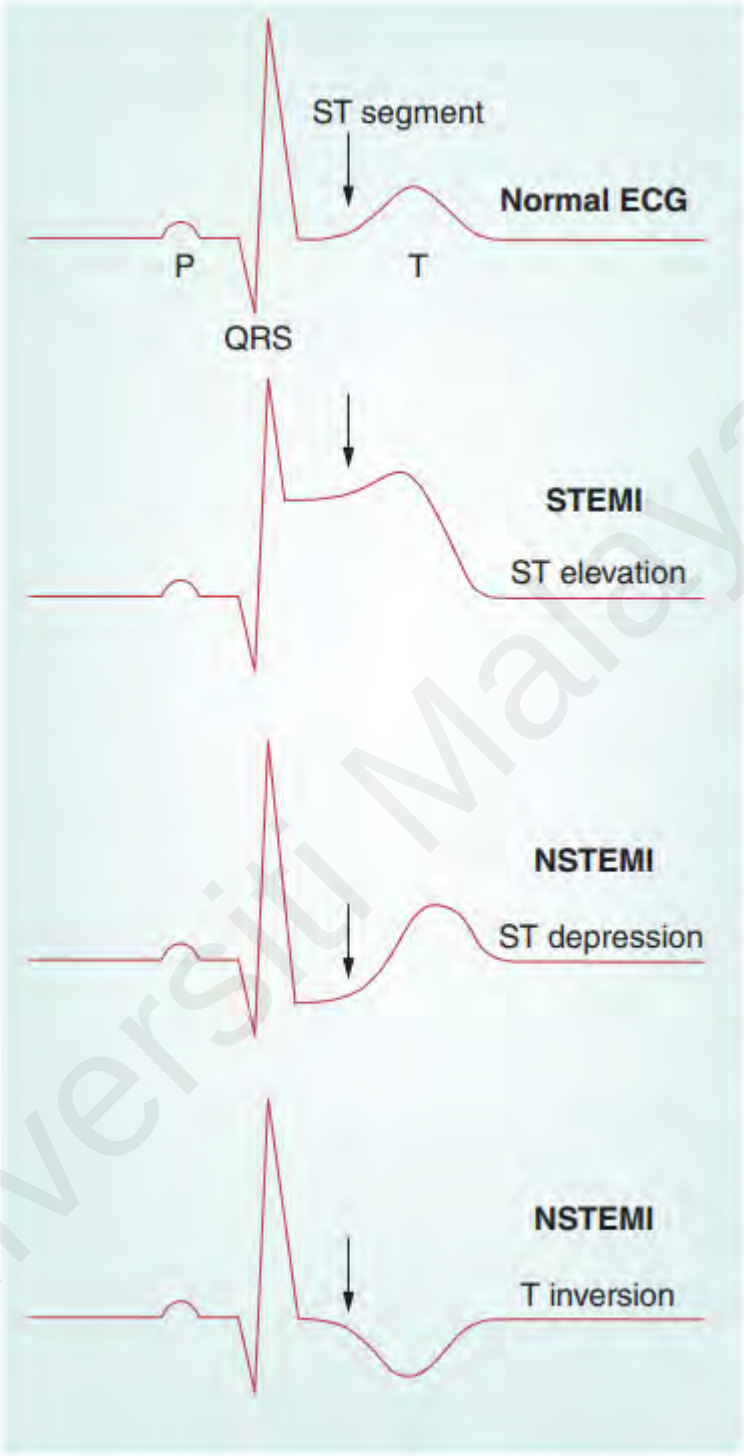


Figure 2.2: The depiction of ECG pattern of normal, STEMI, NSTEMI (ST depression), and NSTEMI (T inversion). (Pleister et al., 2013).

A STEMI heart attack occurs when a blood clot develops abruptly and totally blocks a heart artery. This can cause damage to the heart muscle that reaches deep into it and covers a huge portion of the heart. The goal of STEMI heart attack treatment is to open the artery as rapidly as possible, preserving as much heart muscle as feasible. PCI, which includes both angioplasty and stenting, as well as clot-busting medicines and CABG, are all alternatives for treatment (MOH, 2019).

NSTEMI/UA heart attacks differ from STEMI heart attacks in several ways, the most notable of which is how they appear on an ECG. Damage from a heart attack caused by NSTEMI/UA does not always reach the whole depth of the heart muscle. Different forms of blood clots, with varying quantities of clotting proteins and platelet blood cells, cause NSTEMI/UA heart attacks, much as STEMI heart attacks. As a result, treatment for NSTEMI/UA heart attacks differs from that for STEMI heart attacks. Clot-busting drugs, for example, are inefficient, and while PCI may be utilised as part of the treatment, the primary objective is not to open the artery in less than 90 minutes. As a result, medications to protect the heart and reduce its workload (beta-blockers, nitroglycerin, and possibly an angiotensin-converting-enzyme (ACE) inhibitor or angiotensin-receptor blocker) will be continued, and the patient will only be transferred for coronary angiography with the goal of revascularization if the condition worsens (MOH, 2011).

UA and NSTEMI elevation myocardial infarction are frequently accompanied with a white, platelet-rich, and only partly occlusive thrombus. Microthrombi can detach and embolize downstream, resulting in MI and ischaemia. STEMI, on the other hand, has a more stable occlusive thrombus that is red and fibrin rich.

NSTEMI and UA are loosely related conditions with similar pathophysiologic origins and clinical manifestations, albeit their severity differences (Reed et al., 2017). Due to the

clinical entities in the acute coronary syndrome spectrum of disease overlap, many trials include participants with either UA or NSTEMI, allowing for similar treatment options (Grech & Ramsdale, 2003). Following that statement, in this ACS subtype mortality prediction study, NSTEMI and UA were placed in the same patient category and STEMI in its stand-alone category.

2.2 Risk factor of ACS

Risk stratification and identification are crucial in lowering CVD burden by gathering information from patients about pain features and symptoms, risk factors or a history of cardiovascular disease, and recent medication usage. Risk factors for MI include demographic factors, patient's status before the event, clinical representation, baseline investigation, electrocardiography, treatments, and pharmacological therapy (Torpy et al., 2009).

CVD burden and mortality are also unequally distributed in the population, with substantial differences by age, sex, and race or ethnicity, which is why these variables are critical for predicting mortality (Mensah et al., 2005). Aging can cause changes in the heart and blood vessels that cause high blood pressure, or hypertension (Ahlgren et al., 1997). The difference in gender plays an important role. Men tend to suffer from a heart attack at an earlier age compared to women. Women, however, have a worse outcome compared to men and often present with atypical chest pain and take a longer time to seek medical care (Khesroh et al., 2017). Ethnicity, particularly South-East Asians, has a unique cardiovascular risk profile, with high rates of insulin resistance, glucose intolerance, central obesity, and diabetes, as well as elevated blood levels of other CAD risk factors (Chaturvedi, 2003). In Malaysia, Malays have a higher body mass index (BMI), whilst

Chinese have a greater incidence of hypertension and hyperlipidemia, and Indians have a higher prevalence of diabetes mellitus and a family history of early CAD (Lu & Nordin, 2013).

According to the National Health and Morbidity Survey (NHMS), Malaysia's top morbidity diseases that lead to cardiovascular disease are diabetes mellitus, hypertension, hypercholesterolemia, and smoking (IPH, 2015). Hence, patients' representations prior to admission are also crucial for mortality determination (IPH, 2015). This is because a high level of fats or glucose in the blood causes blood vessels to constrict, disrupting the flow of oxygen to the heart muscles, and ultimately resulting in MI. Smoking, on the other hand, may also contribute to vessel hardening including calcification of the artery wall.

Patients' clinical presentation, such as systolic and diastolic blood pressure and Killip class is essential. Intra-cavitary pressures and the shear stress force of muscle contraction against an inert and necrotic region are affected by blood pressure, resulting in laceration and finally rupture (Birnbaum et al., 2003). The Killip class predicts survival in individuals who have had an acute heart attack, with a higher class indicating a greater risk of death (Juhan et al., 2019).

Lipid profile and glucose level in blood are essential for the prediction of mortality since these are the indicators of the common morbidity disease such as diabetes and hypertension. The observed decrease in cardiac events after lipid treatment is linked to changes in both LDL and HDL levels (Meeusen et al., 2017). Hence it is important for risk stratification. Additionally, there is growing evidence that glucose parameters are independent CVD risk factors (Hryhoriy, 2016).

The electrophysiology of the heart is measured by an ECG. ECG abnormalities reveal the myocardium's electrical instability; hence, ECG can be utilised to screen for susceptible myocardium that could lead to acute MI (Myers et al., 2017). Quantitative metrics derived from resting ECG, ambulatory ECG, and stress ECG have been used to predict CVD events and death in the past.

In stable CAD, the goal of therapy is to improve symptoms and survival. This mostly entails lifestyle adjustments and the most appropriate medical treatment. In patients with left main stem stenosis or complicated three or more vessel coronary artery disease, CABG has been shown to have a survival benefit over PCI, particularly in diabetic patients (Mohr et al., 2013). Angioplasties, also known as PCI, are non-surgical procedures that involve catheter to insert a tiny device called a stent into the heart to widen blood arteries that have been restricted due to the buildup of atherosclerotic plaque while CABG bypasses the restricted or blocked coronary arteries by connecting blood vessels from another section of the body to blood vessels above and below the affected artery (Levine et al., 2016).

ACE inhibitors, beta-blockers, diuretics, calcium channels blockers, Angiotensin receptor blockers (ARB), and antiarrhythmic agents are important to ACS patients because they relieve symptoms, slow disease progression, prevent hospitalisation, and, most importantly, reduce mortality, which is why medications are taken into account when predicting patients' mortality (Packer, 2017). Diabetic patients with cardiovascular risk are also treated with a statin, lipid-lowering agent, and insulin which can dramatically reduce the morbidity and mortality associated with ischaemia, heart failure and control the blood glucose and cholesterol level of patients with diabetes or hypertension (Tran et al., 2020). Table 2.2 summarises prior studies from around the world on the risk factors for ACS.

Table 2.2: List of previous studies conducted in finding the risk predictors that are associated with ACS.

Authors	The population studied (location)	Number of populations	Risk factor finding
(Ralapanawa et al., 2019)	Sri Lanka	300	Smoking, alcohol consumption, hypertension, diabetes miletus, history of ACS, and dyslipidemia
(Borrayo-Sanchez et al., 2018)	Mexico	21,827	Age, gender, BMI, smoking status, hypertension, diabetes mellitus, dyslipidemia, metabolic syndrome (VCEPATPIII), cardiovascular history (previous myocardial infarction, previous angina, cerebrovascular accident), clinical manifestations (typical chest pain, dyspnea, syncope, neurovegetative symptoms) and hemodynamic assessment (heart rate, respiration rate, systolic blood pressure, and diastolic blood pressure)
(Szabó et al., 2021)	Hungary	287	Time to system onset, door to balloon time, age, gender, area at risk, resuscitation, smoking, diabetes, peak creatine kinase level, and hemoglobin.
(Sugane et al., 2021)	Japan	657	Hypertension, chronic kidney disease, maintenance hemodialysis, and history of PCI.
(Vernon et al., 2019)	Australia	3081	Hypertension, diabetes mellitus, hypercholesterolemia, smoking, Killip class, cardiac arrest at admission, systolic blood pressure, and hospital transfer
(Burazeri et al., 2007)	Albania	467	Family history of coronary heart disease, waist-to-hip ratio, hypertension, and smoking status

Table 2.2, continued.

Authors	The population studied (location)	Number of populations	Risk factor finding
(Yadav et al., 2010)	India	200	Smoking status, hypertension, diabetes, family history of CHD, obesity, and dyslipidemia.
(Singh et al., 2020)	India	240	Age, height, weight, BMI, diabetes mellitus, hypertension, dyslipidemia, Medication history, level of exercise, diet, tobacco use, and hospitalization detail.
(Juhan et al., 2019)	Malaysia	16,673	Diabetes mellitus, hypertension, family history of CVD, renal disease, PCI, Killip class, and age
(Ponniah et al., 2012)	Malaysia	603	Age, history of diabetes mellitus, peripheral vascular, renal failure, and previous percutaneous coronary intervention
(Ahmad et al., 2011)	Malaysia	525	Hypertension, diabetes, dyslipidemia, smoking history, previous history of CAD, family history of CAD

2.3 Mortality prediction

When patients are brought to hospitals, a quick decision must be made to avoid any casualties. However, the choice of intervention, treatment plan, and resource allocation must all be considered, and during the last few decades, several general multipurpose mortality assessment systems have been developed to meet these economic and therapeutic objectives. Hence, mortality prediction can be extremely useful in the medical field since it

can help with disease prevention, early detection, and more successful treatment (Lee et al., 2015).

Methods for mortality prediction can be divided into three ways, according to Stoeldraijer et al. (2013): extrapolation, explanation, and expectation.

- **Extrapolative** methods imply that future trends will largely mirror those of the past. This is usually a fair assumption in mortality forecasting because of historical regularities.
- When the important external determinants are known and can be measured, the **explanation** technique uses structural or epidemiological models of mortality from specific causes of death. As a result, critical medical knowledge and data on behavioural and environmental change are put to good use.
- The **expectation** technique is based on the subjective opinions of experts, which can be formal or informal. It's worth mentioning that some mortality prediction systems combine aspects from a variety of methodologies.

Since expert expectations are usually cautious, the expectation approach is not a suitable basis for mortality projection. The same may be said for the explanation technique, which is limited to established causes of death. As a result, rather than models built largely for age-specific validation, most developments have been in extrapolative methods that employ statistical techniques. This strategy takes data from previous patients with the same condition and repurposes it to forecast the outcomes of future patients. Medical claims data is vital for portraying patient health care access regularity and involvement in illness treatment or prevention, which both have a significant influence on patient health outcomes (Tran et al., 2021).

This method is frequently employed in the construction of prediction models. Prediction models represent the distribution of outcomes among people who share a set of traits (Maley et al., 2020). Clinical practice has evolved through time to incorporate the development of predictive models to predict the severity of patient problems and the outcome of treatment measures. Predictive modeling can be thought of as a subset of concurrent analytics, which combines two or more forms of statistical analysis at once. Statistical analyses of huge datasets utilising multivariate risk factor models are typically used to develop such a tool (Mohammed et al., 2014). These assumptions are considered by predictive models in the health industry, which examine patient preferences, demographics, lifestyles, and psychographics.

Over the last two decades, several prediction models have been developed that statistically combine multiple variables to assess the probability of having CVD. These models are also being used to anticipate future cardiovascular disease deaths at the population level and in specific subgroups, in order to provide policymakers and health authorities with information about these risks. Some of these prediction models are recommended by health policymakers and are included in therapeutic management clinical guidelines (Goff et al., 2014). Several studies have found that there is a range of prediction models for various CVD outcomes (Beswick et al., 2011; Matheny et al., 2011; Wessler et al., 2015). According to more recent assessments, the number of published prediction models has risen substantially since then.

The mortality prediction models (commonly known as risk scores) for patients with CVD are discussed further in subchapter 2.4 below.

2.4 Risk Score

Risk scores correspond to multivariate models used in clinical practice to estimate the individual probability of unwanted outcomes. Risk estimates have implications for clinical management, particularly with regard to broad-spectrum diseases, such as ACS (Correia et al., 2014).

This issue is normally happening in the emergency department (ED) and it is not only to identify patients who are at the greatest risk but also to identify patients who have non-urgent conditions or even no disease at all. Patients at low risk for ACS could be identified earlier in the diagnostic process, which could lower patient burden, duration of stay in the ED, hospitalisation frequency, and expenditures (Six et al., 2012).

Normal values of troponin and a normal ECG still do not exclude ACS completely. As a result, many patients presenting with chest pain are currently hospitalized and extensively evaluated with non-invasive stress testing or imaging, or with an invasive coronary angiography (Poldervaart et al., 2017). Although specific demographic and clinical features may be linked to an elevated risk of a negative outcome, the capacity to effectively quantify risk requires the consideration of multiple factors at the same time (Khera et al., 2021).

To be practical clinically, a risk stratification tool should be simple and easily applied at the bedside and should make use of clinical data that are routinely available at hospital presentations. However, to perform accurately, the tool should use data that offer independent prognostic information and must consider the complex profile of patients with multiple risk factors (Correia et al., 2014).

National cardiac guidelines state that chest pain patients presenting to the ED should be assessed with a risk stratification tool or risk score (Damman et al., 2017; MOH, 2011,

2019; Thygesen et al., 2012). These regression-based conventional risk scores were created based on expert opinion and included criteria that the expert considered were more important to clinically diagnose patients by converting prognostic indicators into risk indices. Below is the list of risk scores that have been developed over the years for patients with chest pain in ED.

2.4.1 Thrombolysis in Myocardial Infarction (TIMI) Score

TIMI risk has different models developed for the ACS subtypes; STEMI and NSTEMI, which were validated in distinct samples.

Morrow et al. (2000) established the TIMI risk score for STEMI from the Intravenous nPA for Treatment Infarcted Myocardium Early II trial to predict the mortality of STEMI patients at 30 days. There are eight variables that predict death, each of which contributes points to the scoring when added together. 65 to 74 years old, above 75 years old, diabetes, hypertension, or angina history, systolic blood pressure, heart rate, Killip class, weight, ST-segment elevation in the anterior wall or left bundle branch block, and reperfusion time are the variables. The point for each variable is shown in Table 2.3 below.

Table 2.3: The variables and points for TIMI for STEMI.

TIMI Score Variables	Point
Age between 65-74 years old	2
Age ≥ 75 years old	3
History of diabetes, hypertension, or angina	1
Systolic blood pressure < 100 mmHg	3

Table 2.3, continued.

TIMI Score Variables	Point
Killip classification II to IV	2
Heart rate > 100 bpm	2
Weight < 67 Kg	1
ST-segment elevation in the anterior wall or left bundle branch block	1
Reperfusion time > 4 hours	1

The score will vary from 0 to 14 based on the summation of all the variables presented with the patients at admission. TIMI score 0 to 2 as low, 3 to 5 as intermediate, and >5 as high (Correia et al., 2014).

Antman et al. (2000) used the TIMI 11B clinical trial for the composite endpoint of mortality at 14 days to build the TIMI risk score for NSTEMI/UA in the year 2000. This risk score is used to help patients with suspected ischemic chest pain, usually those with NSTEMI/UA, risk stratify. Age >65 years, 3 classical risk factors for coronary artery disease (CAD), known CAD (stenosis >50%), use of aspirin in the previous 7 days, severe angina in the previous 24 hours, elevated cardiac markers, and ST-deviation 0.5 mm are the 7 dichotomous variables that made up the scores. Each variable is assigned a point value of 0 or 1, and the total score will range from 0 to 7. Patients with a score of 0 to 2 points are deemed low risk, intermediate risk at 3-4 points, and high risk at 5-7 points. The point for each of the variables for patients with NSTEMI/UA is shown in Table 2.4 below.

Table 2.4: The list of variables and points for the TIMI risk score for NSTEMI/UA.

TIMI Score Variables	Point
Age more than 65 years	1
At least 3 risk factors for CAD (family history of premature CAD, hypertension, elevated cholesterols, active smoker, diabetes)	1
Known CAD (coronary stenosis of > 50%)	1
Aspirin use in within 7 days	1
ST-segment deviation (>0.5mm) on ECG	1
At least 2 anginal episodes in prior 24h	1
Elevated serum cardiac biomarkers	1

The TIMI risk score, according to Morrow et al. (2000), is ideal for developing countries since it allows for low-cost risk estimates. It was created in the context of a clinical trial. However, it was primarily taken from a Western cohort, with non-western participants contributing less. It was able to accurately predict 30-day and one-year mortality. In their analysis, Feder et al. (2015) identified several of the TIMI risk score's strengths, including its widespread familiarity among medical professionals, ease of use, and reliability, as demonstrated by a vast evidence base of development and validation studies. The TIMI risk score has limitations, including those inherent in the trial score and the exclusion of high-

risk patients. While the lack of risk factor weighting improved usability, it reduced discriminatory performance.

Despite the limitations, the simplicity of the TIMI score is recognized in the current guidelines. It has also been used in key studies to demonstrate the benefit of clopidogrel at all risk levels and to demonstrate graded benefits of tirofiban with increasing risk levels.

2.4.2 Global Registry of Acute Coronary Events (GRACE) Score

The GRACE score was developed in the year 2003 by Granger et al. (2003) from a multinational registry of 11, 398 patients ACS patients. Data registration was done prospectively and retrospectively. The GRACE score is based on participants in a registry who did not get any experimental treatment. Patients in this registry, on the other hand, had to have a definitive diagnosis of ACS and were only included if they exhibited ECG abnormalities indicating ACS, a sequential increase in cardiac enzymes, or confirmed CAD. Then, multivariate LR was adopted in building the risk score which is used for the prediction of in-hospital and post-discharge death at 6 months. There are 8 variables for the mortality prediction of patients in in-hospital which are Killip class, age, blood pressure, resuscitated cardiac arrest, positive cardiac marker findings, creatinine level, ST-segment deviation, and heart rate. While the variables for the 6-month post-discharge mortality prediction are age, congestive heart failure, MI, heart rate, systolic blood pressure, ST-segment depression, serum creatinine, elevated cardiac markers, and no in-hospital PCI. The comparison between the two prediction time points can be seen in Table 2.5 below. The GRACE in-hospital risk score (range 0–372) and GRACE 6-month risk score (range 0–263) were developed from the GRACE registry for the endpoint of all-cause death and consist of eight and nine factors respectively.

Table 2.5: The GRACE score comparison of variables between the two points prediction.

GRACE score for in-hospital mortality	GRACE score for post-discharge 6 months mortality
Age	Age
Heart rate	H/o congestive heart failure
Systolic blood pressure	H/o myocardial infarction
Serum creatinine level	Heart rate
Killip class	Systolic blood pressure
Cardiac arrest at admission	ST-segment depression
Elevated cardiac markers	Serum Creatinine
ST-segment deviation	Elevated cardiac marker
	No in-hospital PCI

According to the research done by Shuvy et al. (2018), due to its capacity to stratify patients, the GRACE risk score considerably lowered the death rate of ACS patients. As a result, high-risk individuals were able to obtain medical care and preventable treatment at the right moment. According to Khalill et al. (2009), GRACE had a better performance compared with other risk scores because it is a powerful predictor to calculate the risk more precisely and the associated mortality rate. Moreover, GRACE is easier to conduct and use.

According to Bassand et al. (2007), in the prediction of in-hospital mortality rate, ACS patients are at low risk when the GRACE risk score is lower or equal to 108, with the probability of in-hospital death of lower than 1 %. The patients are at intermediate risk if the GRACE risk score falls between 109 to 140, with the probability of in-hospital death of 1 to 3%. However, the patients who score more than 140 will fall into the high-risk category with in-hospital death of more than 3%.

On the other hand, in the prediction of post-discharge to 6 months mortality rate, ACS patients are at low risk when the GRACE risk score is lower or equal to 88, with the probability of post-discharge to 6 months death of lower than 3%. The patients are at intermediate risk if the GRACE risk score falls between 89 to 118, with the post-discharge to 6 months probability of death of 3% to 8%. ACS patients who score more than 118 will fall into the high-risk category post-discharge to 6 months probability of death with of more than 8%.

2.4.3 Platelet glycoprotein IIb/IIIa in Unstable angina: Receptor Suppression Using Integrilin (eptifibatide) Therapy (PURSUIT) Score

The PURSUIT score was developed in the year 2000 in a multinational randomized clinical trial with 9,461 patients (Platelet glycoprotein IIb/IIIa in Unstable Angina: Receptor Suppression Using Integrilin (eptifibatide) Therapy) comparing eptifibatide (Integrilin) to placebo in the management of UA or NSTEMI (Boersma et al., 2000). This population is representative of a wide range of patients, hospital settings, and treatment policies, making it ideal for the creation of a clinical risk model.

Using multivariate regression analysis, the researchers revealed seven risk factors for mortality and MI in ACS patients. Higher age, gender, the worst Canadian Cardiovascular Society class of angina, heart failure symptoms, and ST-segment depression

on the index ECG were all factored into the scoring system. The researchers did not consider tachycardia or low systolic blood pressure in their final risk score (Table 2.6) (Investigators, 1998). The PURSUIT score predicts the risk of death or MI at 30 days after admission.

Table 2.6: List of variables for PURSUIT score prediction.

PURSUIT variables		Points
Age (decade)	50	8
	60	9
	70	11
	80	12
Sex	Male	1
	Female	0
Worst CCS class passed 6 weeks	No angina/CCS I, II	0
	CCS III/IV	2
Signs of heart failure		2
ST depression on ECG		1

According to the PURSUIT score, ACS patients are classified as low (0 to 5 points), intermediate (6 to 9), or high risk (10 and above), “with the early release”, “watchful waiting,” and “aggressive antiplatelet/early invasive procedures as proposed therapy”, respectively (de Araújo Gonçalves et al., 2005).

PURSUIT score is giving a moderate result in predicting in-hospital mortality, all-cause mortality, and 1-year follow-up all occurrence or non-ACS occurrence mortality with AUC more than 0.7.

2.4.4 HEART Score

The HEART risk score was created to stratify patients who presented to the emergency department with suspected NSTEMI (Six et al., 2008). Patients with a higher risk of MACE (all-cause death, myocardial infarction, or coronary revascularization) in the next 6 weeks are identified in this study. The HEART score was created using decision-making clinical factors rather than multivariate regression analysis, according to expert judgment. The European Society of Cardiology created this to anticipate ACS in order to improve health and reduce risks in people with cardiovascular issues. The HEART score is made up of five clinical judgment parameters: history, ECG, age, risk factors, and troponin to which its acronym makes up the name of it – HEART. The HEART score was verified in 122 patients in a single centre retrospective study and 880 participants in a multicenter retrospective analysis (Backus et al., 2010). Each of the variables in the score will have given 0,1 or 2 points which its summation will range from 3 to 10 (Table 2.7).

Table 2.7: The variables and their attributes in the HEART Score.

HEART Score variables		Points
History	Highly suspicious	3
	Moderately suspicious	2
	Slightly or non-suspicious	1

Table 2.7, continued.

HEART Score variables		Points
ECG	Significant ST-depression	3
	Non-specific repolarization	2
	Normal	1
Age	≥65	3
	46-64	2
	≤45	1
Risk factors (Diabetes, current or recent (less than one month) smoker, hypertension, hypercholesterolemia, family history of CAD and obesity)	≥ 3 risk factors or history of CAD	3
	1 or 2 risk factor(s)	2
	No risk factor	1
Troponin	≥ 3x normal limit	3
	> than 1 to < than 3 normal limits	2
	≤ normal limit	1

The HEART score categorises individuals into three risk groups: low (0-3), intermediate (4-6), and high (7-10), with mean event risks of 0.9%, 12%, and 65%, respectively. As a consequence, an evidence-based decision on whether the patient should be released from the ED (low-risk patient), hospitalised for clinical observation (intermediate-risk patient), or immediately treated with invasive therapy (high-risk patient) may be made. Despite the fact that the authors do not suggest that patients with a HEART score of 3 or below should be safely discharged without further examination, they do claim that the HEART score can be used to "triage" patients with chest discomfort because it is a "reliable predictor of prognosis" (Fesmire et al., 2012).

2.4.5 Fast Revascularization in Instability in Coronary Disease (FRISC) Score

The FRISC score (Fast Revascularization in Instability in Coronary Disease) II study is the basis for the FRISC score (Lagerqvist et al., 2005). A multicenter, randomised clinical study including individuals with unstable CAD in Scandinavia was used to establish this risk score. Using multivariate regression analysis, data from 1,235 patients in the non-invasive cohort were used to identify seven factors as independent predictors of 1-year mortality or death/MI in patients with UA. The FRISC score is made up of seven parameters: age 70, male gender, diabetes, previous MI, ST-segment depression on admission, increased Troponin, and elevated Interleukin 6 or CRP levels. Each of these elements is worth 0 or 1 point, giving in a total score of 0-7 (Table 2.8).

Table 2.8: The variables in the FRISC score model.

FRISC Score variables	Points
Age ≥70 years	0
	1

Table 2.8, continued.

FRISC Score variables	Points
Male sex	0
	1
Diabetes	0
	1
Previous MI	0
	1
ST-depression on ECG	0
	1
Elevated Troponin levels	0
	1
Elevated Interleukin 6 or CRP	0
	1

The FRISC scores of 0-2, 3-4, and 5-7 were used to divide patients into low, middle, and high-risk groups, respectively. According to Lagerqvist et al. (2005), the FRISC score, which is based on seven criteria, is widely available in UA and is extremely effective for risk classification and selecting patients for an early invasive procedure. Patients who have three or more of these criteria should be treated with an early invasive strategy, but those

who have 0–2 criteria have low event rates regardless of the treatment method. The FRISC score seems to be the only risk assessment in ACS that focuses on the early invasive approaches' treatment effect.

Table 2.9: The common conventional risk scores for heart risk prediction.

Variables	Conventional risk score					
	TIMI for STEMI	TIMI for NSTEMI/UA	GRACE	PURSUIT	HEART	FRISC
Age	•	•	•	•	•	•
Sex				•		•
Past medical history	•	•			•	•
Risk factors		•			•	
Medication used		•				
CSS/Killip class	•		•	•		
Signs and symptoms		•		•		
Patient history					•	
Cardiac arrest upon admission			•			
Heart rate	•		•			
Systolic blood pressure	•		•			
Weight	•					

Table 2.9, continued.

Variables	Conventional risk score					
	TIMI for STEMI	TIMI for NSTEMI/UA	GRACE	PURSUIT	HEART	FRISC
ECG findings	•	•	•	•	•	•
Cardiac enzymes		•	•		•	•
Creatinine level			•			
Treatment time	•					
Possible range of scores		0-7	0-372	0-18	0-10	0-5
Cutoff of high risk		≥3	≥109	≥10	≥4	≥2

2.4.6 The limitation of risk scores.

The following are some of the major and significant challenges that ACS patients experienced in earlier CVD prediction algorithms (risk score). To begin with, most previous regression-based CVD prediction models fail to accurately predict and diagnose CVD events in moderate-risk people. Nearly half of MIs and strokes, for example, occur in people who are not thought to have CVD (Ridker et al., 2008). Even though standards for CVD risk diagnosis and prediction are available, doctors frequently treat individuals with intermediate-risk unnecessarily. Secondly, conventional regression-based CVD prediction algorithms contain common and frequently used prognostic parameters such as age, blood pressure, heart rate, diabetes, cholesterol, smoking, and heart disease history and do not introduce different prognostic factors that might assist in the prediction of the desired

outcome as certain risk factor combinations may work together synergistically to raise risk in a way that is more than additive (Cooney et al., 2009). Additionally, these risk scores were derived from the Western Cohort with only limited numbers of participants. Hence, the risk scores may not reflect the region's diversity and maybe only applicable to specific populations (Peng et al., 2017). Although these models have been validated and are widely used, there have been recent concerns expressed because most traditional risk stratifications were built 20 years ago using randomised controlled trials (RCT) data prior to the introduction of drug-eluting stents and newer generation antiplatelets (Kwon et al., 2019a). Furthermore, the prediction models' outcomes are limited to short-term mortality, such as mortality in the hospital, 14-day mortality, and 30-day mortality. As a result, according to one review of traditional risk stratification models, future models will allow for more exact risk stratification (Castro-Dominguez et al., 2018). Finally, the non-linear interactive interactions among prognostic factors are oversimplified because each prognostic factor in the regression-based CVD prediction model is connected to the incidence of major cardiovascular events, which are identified as a composite of death, MI, or repeat coronary revascularization of the target lesion (Ahmed & Hannan, 2012). As a result, models including these various risk variables and outcomes, as well as the usage of AI algorithms, are required (Obermeyer & Emanuel, 2016; Peng et al., 2017).

2.5 Artificial intelligence (AI)

The use of AI algorithms to discover patterns from large data sets in order to better predict mortality is a recent trend. (Booth & Tickle, 2008). Historical electronic health records (EHRs) are widely utilised to build AI models that predict patient health outcomes, according to the medical literature. Patient demographics, health indices, health problems,

biological images, and patient records are all common data sources for AI models, however, organised medical claims data is rarely used (Tran et al., 2021).

As it can handle and optimise very complicated datasets existing in very complex systems, true AI has a lot of applications in health care. Taking care of patients requires the control of numerous processes, each of which is extremely variable and dependent on or connected to multiple other steps. While keeping track of so many variables is difficult for humans, computers are exceptionally adept at it (Bini, 2018).

The AI techniques are recognised as ways to get beyond the limitations of standard CVD incidence prediction models, which are used to generate traditional risk scores. By making precision cardiovascular investigations simpler, AI approaches like ML and DL may play a crucial role in the evolution of cardiovascular medicine (Krittanawong et al., 2017).

Table 2.10: Comparison of conventional risk prediction and AI-based risk prediction approach.

Feature	Conventional risk score	AI-based risk prediction
Hypothesis	Yes	No
Approach	Estimates and explain data	Practical prediction from data
Measurement	Goodness-of-fit, coefficients	Precision, recall, F-measure, accuracy, area under the curve
Learning ability	No	Yes
Data size	A proper data size for a certain hypothesis	Big and complex data
Data type	A single type of data, structured data	Multi-modality data, structured and unstructured data are all supported.
Model	Simple parametric model	Complex, non-parametric model

Table 2.10, continued.

Feature	Conventional risk score	AI-based risk prediction
Output	Validate the hypothesis, causality	Predict new data, identify new patterns
Limitation	Low data dimensionality and require assumptions	Overfitting, data privacy, and security issues
Risk factors	Clinical or demographic factors only e.g. age, gender, smoking, diabetes	Multimodality e.g. age, gender, ECG variations, treatment, features from an image, gene expressions

2.5.1 Machine learning (ML)

According to Bini (2018), the best way to think of ML is as a subset of AI. ML is a discipline in computer science that uses a range of computational algorithms to allow computers to improve their performance on a particular task incrementally rather than being explicitly programmed (Pieszko et al., 2019). Instead of having to predetermine the system's mechanical linkages, which could yield more knowledge and information, ML techniques use observed data to "learn" information about a system (Li et al., 2020). To create a decent classification and prediction, ML learns from its experiences and improves its performance over time by recognising a given pattern from training data (Bini, 2018).

Gibson et al. (2020) in their study stated that ML-based models were able to map extremely non-linear input and output patterns even when mechanistic relationships between model variables could not be found due to pathologies or complexity. Not only can ML algorithms detect interaction, nonlinear, and higher-order effects, but they can also estimate complex functions that are not properly represented by a single covariate or interaction term (Al'Aref et al., 2019). For its ability to modify performance with each new

data sample, ML has become the go-to technology for a range of real-world applications (Shaikhina et al., 2015). Different ML algorithms may be required for an application, according to Libbrecht and Noble (2015), depending on whether one is interested in interpreting the output model or only concerned with predictive power.

ML is neatly divided into two categories: supervised learning and unsupervised learning (Rajkomar et al., 2019). According to Panch et al. (2018), supervised learning is the process of training computer systems to learn the relationships between data inputs and outcomes. After learning the correlations, they can be used to anticipate future examples based on current data. Supervised learning is also known as the development of algorithms that use externally given cases to predict the fate of future instances by establishing broad patterns and hypotheses (Amanpreet et al., 2016). It focuses on classification and prediction, which requires choosing among subgroups to effectively describe a new instance of data (Deo, 2015). Predictive models for medical diagnosis have been built using supervised ML techniques (Maroco et al., 2011). Some of the commonly used ML methods in the medical field are logistic regression (LR), support vector machine (SVM), random forest (RF), artificial neural network (ANN), k-nearest neighbor (kNN), Naïve Base (NB), and decision tree (DT) (Chandralekha & Shenbagavadivu, 2018).

Unsupervised learning, on the other hand, has no outputs to predict and is used to detect naturally occurring patterns or groupings in data (Kohonen et al., 2001). This is a more challenging task to assess, and the utility of unsupervised learning groups is frequently decided by how well they perform in subsequent supervised learning tasks. When the instances are unlabeled, these algorithms attempt to apply techniques to the input data in order to mine for rules, find patterns, summarise, and aggregate the data points, which

assist in extracting useful insight and better conveying the data to the user. Self-Organizing Map (SOM) is a popular unsupervised learning approach.

The following are some of the most commonly utilised ML algorithms in the construction of predictive models which was also used in this study too:

2.5.1.1 Random forest (RF)

RF is an ensemble approach that uses bootstrapping samples to create numerous decision trees, which are then grouped using a classification or regression algorithm (Breiman, 2001). The fact that the RF provides an internal assessment of the relative relevance of each feature on the prediction is a unique feature. This model works well for almost any situation, regardless of size or whether the data is unbalanced or absent (Fernandez-Lozano et al., 2021). Figure 2.3 depicts the RF model development technique.

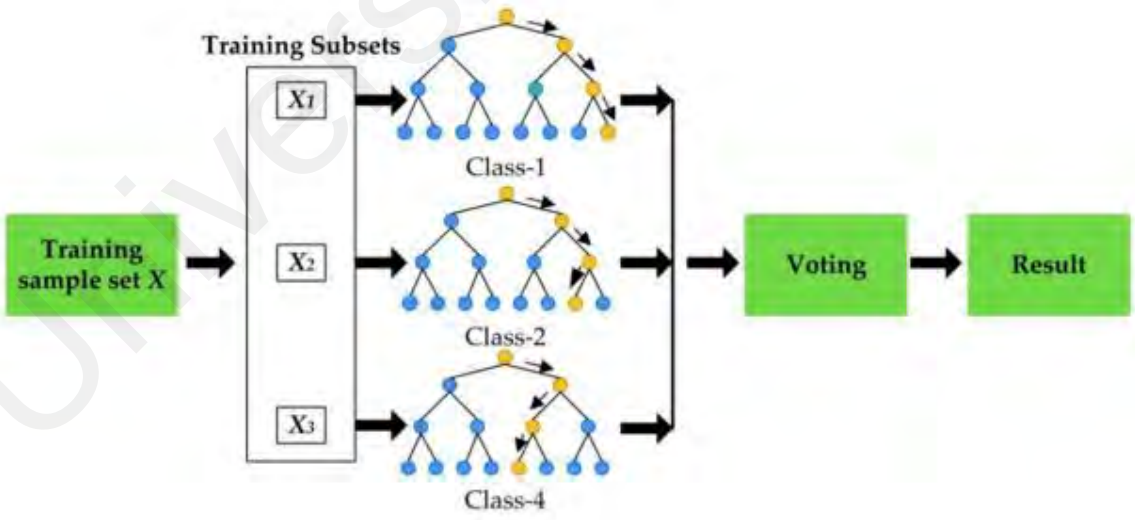


Figure 2.3: The architecture of RF (Wu et al., 2019).

In RF, there is additional randomness added to it (Liaw & Wiener, 2002). The difference between RF and other trees is that in RF, only a subset of predictors is chosen randomly from the full set of predictors, *p*, which is denoted by *mtry*, at each node, and the best split is calculated using Gini index node of impurity, which calculates only among the subset of predictors (Genuer et al., 2010). The Gini index of impurity is a metric for class label distribution at each node. The Gini impurity values are 0 and 1, with 0 indicating that all predictors at the node have the same class history (Khalilia et al., 2011). The smallest Gini impurity value among the predictors is used to make the optimal split choice. To lower the error rate, $mtry=\sqrt{p}$ (for classification) or $mtry=p/3$ at each node of the tree (for regression) is used. Because no pruning is required in RF, the trees formed are maximum, low-bias, and low-correlation (Díaz-Uriarte & De Andres, 2006). RF has several appealing features, including a limited number of tunable parameters, automatic calculation of generalisation errors, and high resistance to overfitting (Wang et al., 2021).

The classification algorithm of the RF is as follows:

1) A bootstrap sample of the training data is used to grow each RF tree.

2) When constructing a tree, *n* number of variables are randomly chosen from *N* predictors at each node.

3) It is proposed that the value of *n* begins at $n=\sqrt{N}$ and grows until the minimum out-of-bag (OOB) error is reached. From all *n* values, one variable with the best split is utilised at each node.

4) The RF model is then used to test data and make predictions.

The two most well-known methods for building the tree are boosting (Schapire & Singer, 1999) and bootstrap aggregating (also known as bagging) (Breiman, 1996) of prediction trees. Boosting operates by producing subset *s1* from the training set without replacement and training *s1*. Then, a new subset *s2* is produced by taking samples from *s1* with the 50% of samples that were misclassified and *s2* is trained. Every subset is dependent on each other, and prediction is made by weighted vote. Bagging is a process of constructing the tree by producing multiple training sets from the original dataset with replacement. The training sets are called the bootstrap samples which then be used to build a model. The bootstrap sample is known to be independent of the original sample hence each of the bootstrap samples votes with equal weight (Liaw & Wiener, 2002). This method is designed to reduce the variance and over-fitting. Thus, most of the base classifiers would consistently detect only truly present patterns in the data and the majority votes turn out to be good class indicators (Amaratunga et al., 2008). The capacity to accommodate larger data inputs, non-linear variables, variable interactions, and minimise overfitting are all advantages of RF models (Kruppa et al., 2012; Peng et al., 2010).

2.5.1.2 Support vector machine (SVM)

The SVM (Vapnik, 1999) has been used as one of the most powerful classifiers for decades because it has proven to outperform other classifiers. The SVM (Cortes & Vapnik, 1995) is a supervised learning model that uses labelled data to learn. It generates a set of labelled input-output mapping functions as well as additional information. The SVM can be used as a classification or regression approach (Orbann et al., 2017).

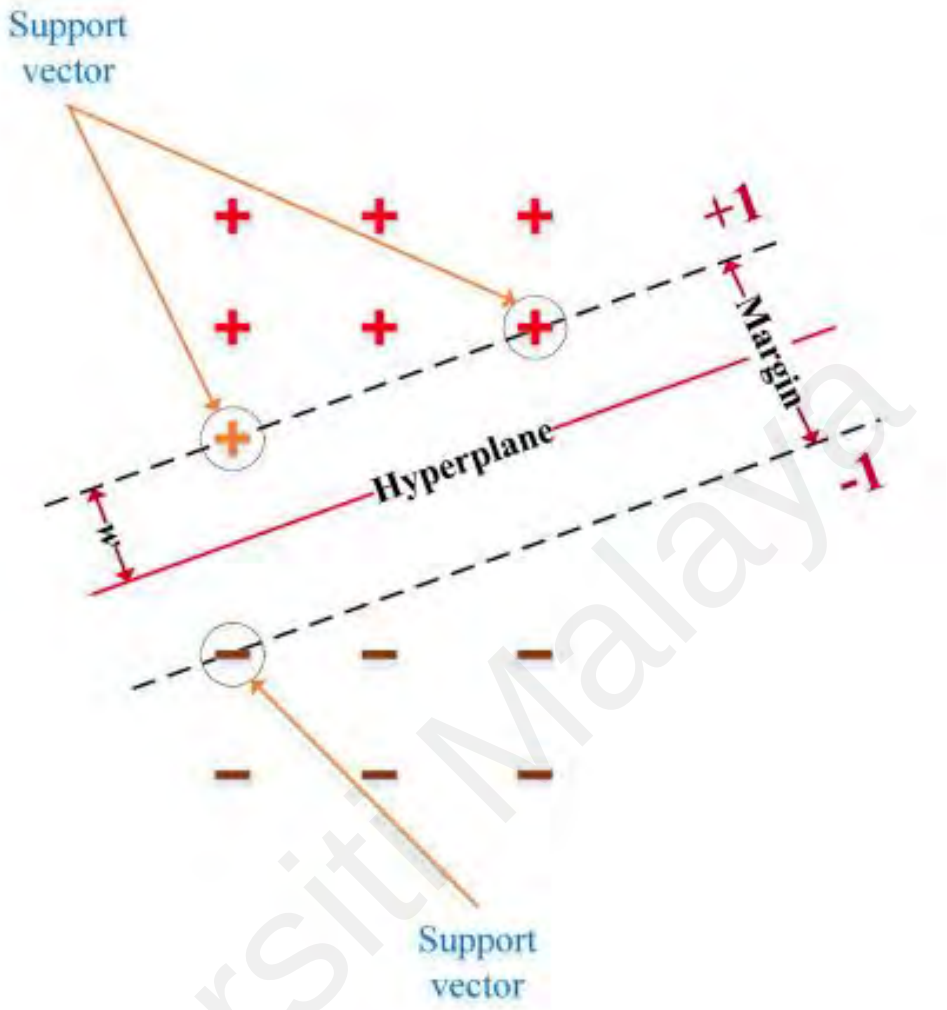


Figure 2.4: The depiction of SVM architecture (Zeng et al., 2021).

When using grid search, which is also known as a big margin classifier, SVM uses optimization parameters. In simple binary scenarios, the two classes divide linearly, and the hyperplane, as shown in Figure 2.4, defines the boundary between the two classes. The SVM generates a classification hyperplane in the middle of the most important margin to identify the ideal group division margin (Marjanović et al., 2011). These two classes are labelled "+1" (positive samples) and "-1" (negative samples), with "+1" denoting a circumstance above the hyperplane and "-1" denoting a situation below the hyperplane. The

properties of new data can then be utilised to determine which assortment a new record should be placed in. The Kernel function, which is analytical and precise, is used to transmute the data. The SVM classifier's kernelisation allows the learning to take place in the feature space. The inner product of the pictures of two data points in feature space is returned by the kernel function (Karatzoglou et al., 2006). This is known as the "kernel trick" in literature (Schölkopf et al., 2002).

The kernel function is a mathematical method that allows the SVM to classify a set of originally one-dimensional data into two dimensions. A kernel function, in general, projects data from a low-dimensional space to a higher-dimensional space (Cortes & Vapnik, 1995). Linear, Radial Basic Function (RBF), and Polynomial are some of the most commonly utilised kernels in SVM (Rai & Khanna, 2011).

When the data is linearly separable, that is, when it can be separated using a single line, the Linear Kernel is utilised. It is one of the most often utilised kernels. It is most commonly utilised when a data set contains a large number of features. The linear kernel is a basic kernel function based on the penalty parameter C, which manages the trade-off between error frequencies and decision rule complexity, however, it is not suited for large datasets (Cortes & Vapnik, 1995).

$$\kappa(\chi_\iota, \chi_j) = 1 + \chi_\iota^\tau \chi_j \tag{2.1}$$

Polynomial kernel function, also known as the global kernel, is a non-stochastic kernel estimate using two parameters, *C*, the penalty parameter, and *d*, the degree of the kernel function. The output of the polynomial kernel function depends on the direction of the two vectors in low-dimensional space (Prajapati & Patle, 2010). Regardless of its actual distance from χ_ι, each data from the set *xi* affects the kernel point of the test value χ_j. With

a small number of support vectors and a low classification error, it provides good classification accuracy.

$$\kappa(\chi_\iota, \chi_j) = (1 + \chi_\iota^\tau \chi_j)d \tag{2.2}$$

Boser, et al. proposed complicated relationships in SVM by changing each dot product with different types of non-linear functions (Boser B.E. et al., 1992). One of the most widely used kernel functions is Radial Basis Function (RBF) (Cristianini & Shawe-Taylor, 2000). RBF also known as the local kernel, is equivalent to transforming the data into an infinite-dimensional Hilbert space. As a result, the non-linear classification problem is simply solved. RBF produces similar results as the polynomial with the lowest training error, however the number of support vectors and classification error increase in some circumstances (Rojo-Álvarez et al., 2018).

$$\kappa(\chi_\iota, \chi_j) = exp\left(-\gamma \|\chi_\iota, \chi_j\|^2\right) \tag{2.3}$$

The free variable of Gaussian RBF is gamma, γ; this variable specifies how far a training sample's impact spreads. Gamma, γ variable is the inverse of the radius of the data impact chosen by the model as support vectors. This indicates that high Gamma, γ will only evaluate points near the plausible hyperplane, whereas low Gamma will consider sites further away. The Cost, *C* (penalty parameter) variable balances the decision surface's simplicity with the misclassification of training data. If *C* is higher, the optimization will choose a narrower margin hyperplane, resulting in a reduced rate of training data misclassification. If the C is low, on the other hand, the margin will be large, even if there are some incorrectly classified training data samples (misclassification).

According to Prajapati and Patle (2010), the advantages of SVM are more than two predictor variables can be handled using SVM, non-linear curves are considered to separate

the points, capable of dealing with clusters that can't be separated, able more than two-category categorization and SVM can handle a dataset with a large number of attributes.

2.5.1.3 Logistic regression (LR)

The purpose of LR is to find the best model to represent the relationship between a set of independent variables and a dichotomous characteristic of a dependent variable (Le Cessie & Van Houwelingen, 1992). LR is also known to be multinomial if the outcome takes more than two values. The coefficients of a formula to predict a logit transformation of the probability of the presence of the feature of interest was developed using LR. In other words, LR analysis calculates an event's log odds. LR calculates a multivariate linear regression function mathematically (Juhan et al., 2019).

The logistic transformation of the probability for each class in the dependent variable is predicted using a sigmoid function in this model. The logged odds assign a binary classification to the data points (Hernandez-Suarez et al., 2019). In addition to conjugate gradient descent, the lambda parameter utilised in a model is a ridge value of 1.0E-8. The cost function of the model is reduced via conjugate gradient descent.

$$logit(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \tag{2.4}$$

Based on equation 2.4, p is the probability of the characteristic of interest present, β_1, is the logistic coefficients and x_1, is the independent variable where, i = 1,...,n.

Stepwise regression analysis is commonly used in building a predictive model of logistic regression by finding the variable of importance (Hosmer Jr et al., 2013). It is an approach that allows changing the course by removing or adding variables at each stage. The significance of the score statistic is used to add variables, and the probability of a

likelihood-ratio statistic based on the maximum partial likelihood estimates is used to remove variables. If the p-value is less than 0.05, the variable is considered significant.

2.5.2 Deep Learning (DL)

DL is defined as a sub-class of ML within the AI technologies that explores many layers of non-linear information processing for supervised and/or unsupervised features extraction and transformation, and pattern analyses and classification (Dey et al., 2020; Diez-Olivan et al., 2019). The concept of successive layers of representations is represented by the word "deep" in deep learning. The depth of a model refers to the number of layers that make up the model. Furthermore, DL will automate the feature engineering process, learning all features in a single pass (Chollet, 2018). Figure 2.5 below summarises the relationship between DL, ML and AI.

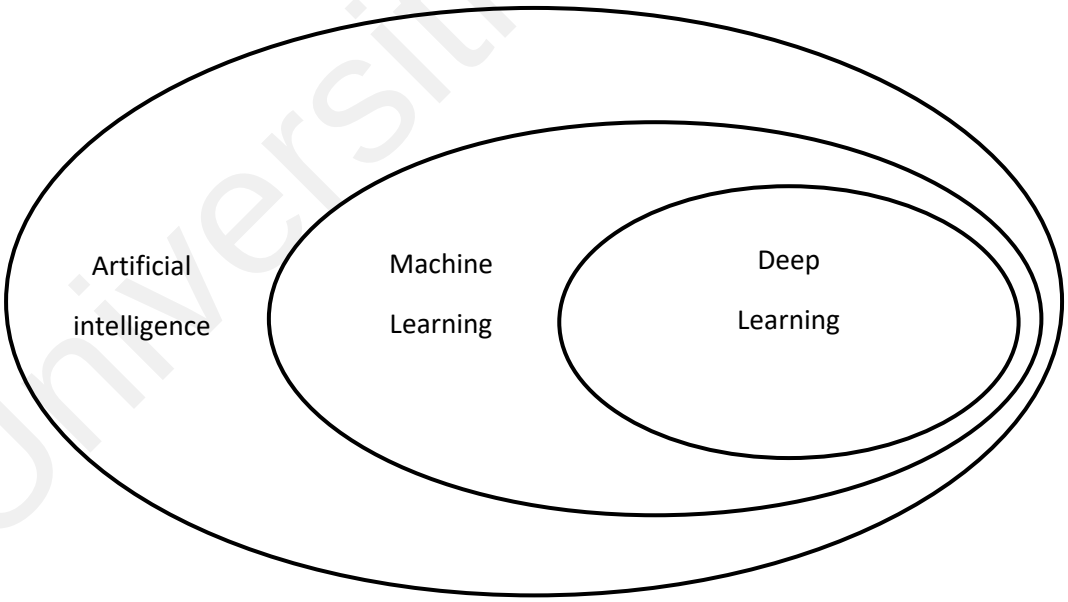


Figure 2.5: The depiction of the relationship between Artificial Intelligence, Machine Learning, and Deep Learning.

DL approaches allow a computer to process enormous amounts of raw data and discover the representations needed for detection or classification. DL methods rely on several layers of data representation with repeated modifications that amplify important characteristics of the input while suppressing unimportant variations (Panch et al., 2018).

Due to the availability of massive data and the development of computational methods that simplify the training of large neural networks, deep neural network learning has resurged considerably in recent years (Schlesinger & Stultz, 2020).

2.5.2.1 Multilayer Perceptron (MLP)

An input layer, many hidden layers, and an output layer make up the basic multilayer perceptron (MLP) structure. MLP is an artificial neural network that works similarly to the human brain in that it learns and stores information in interneuron connection strengths known as synaptic weights (Min et al., 2017). MLPs are one of the most widely used neural network architectures, having been first defined by Rumelhart et al. (1988). MLP is a supervised network because it learns through desired output.

Error backpropagation is one of the most widely used algorithms in MLP training, according to Gurgen et al. (2000). For MLP feedforward networks, it is an optimization technique to apply gradient descent in weight space.

It is trained using only labelled data and is completely supervised. The MLP receives the input data first, and the output values are computed progressively alongside the network layers. The weighted sum is obtained by multiplying the input vector holding the output values of each unit in the layer below by the weight vector for each unit in the current layer. The weighted total is then multiplied by a non-linear function such as a sigmoid, hyperbolic tangent, or rectified linear unit (ReLU) to determine the layer's output values (Min et al.,

2017). It is the nature of neural network models to improve their predictions by iteratively comparing their forecasts to observed results and then adjusting their weight parameters to enhance their predictions (Ramchoun et al., 2016; Sanderson et al., 2019).

Learning for MLP is the process of modifying the connection weights to achieve the smallest difference between the network's output and the desired output. The MLP architecture is determined by the number of layers, the number of hidden neurons in the hidden layers, and the objective functions (Ramchoun et al., 2016). The components of the MLP architecture are depicted in Figure 2.6.

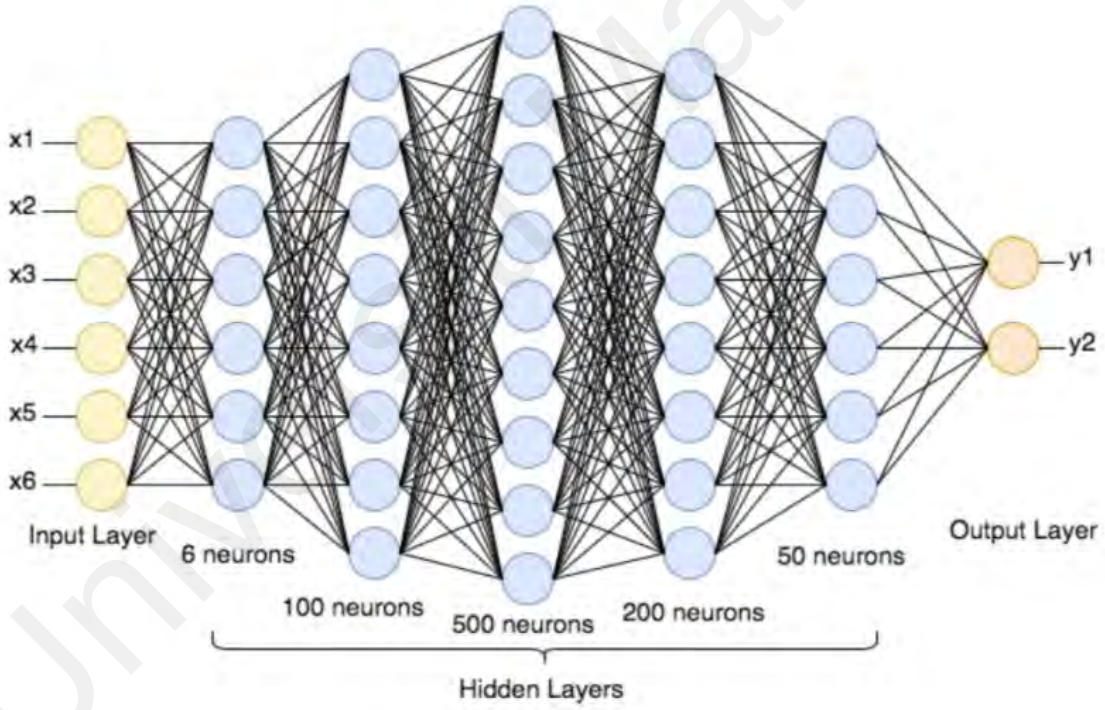


Figure 2.6: The basic structure of MLPs with input units, x1 to x6, five hidden units in each layer, and output units, y1 and y2 (Bahi & Batouche, 2018).

2.5.2.2 Deep Learning Libraries

TensorFlow makes it easier and faster to explore and apply neural network models. It's a more advanced library for distributed numerical computing. Large neural networks can be efficiently trained and run by distributing computations among hundreds of multi-GPU (graphics processing unit) computers. TensorFlow is recommended as the default backend for most DL purposes due to its widespread acceptance, scalability, and the fact that it was mostly production-ready (Chollet, 2018).

Keras is a high-level DL API that simplifies and accelerates the training and running of neural networks. It works with Tensorflow, Theano, and Microsoft Cognitive Toolkit (formerly known as CNTK).

Keras is a DL framework, according to Chollet (2018), that makes it simple to define and train practically any type of DL model. Keras was designed with researchers in mind, allowing for quick experimentation. Keras has several notable characteristics, including a user-friendly API that can support both convolutional and recurrent networks, as well as support for arbitrary network designs.

2.5.2.3 Deep Learning Hyperparameters

Min et al. (2017) state that adopting a suitable DL architecture is critical for obtaining robust and trustworthy outcomes. Choosing a thorough technique for the most appropriate or "best fit" DL architecture, on the other hand, remains a challenge that will be studied in the future.

Many hyperparameters in DL architecture can be tweaked, including the number of hidden layers, hidden neurons, weight initialization values, learning iterations, and learning rate. The setting of these hyperparameters will have a significant impact on the training

outcome. For many years, however, hyperparameter tweaking is rarely systematic, and hyperparameter tuning is still handled by human-ML professionals. The use of automation in hyperparameter tuning is on the rise.

The input data is mapped to the predictions by the network, which is made up of layers that are coupled together. The network's predictions are then compared to the target, yielding a loss value: a measure of how closely the network's predictions match what was predicted. This loss value is used by the optimizer to adjust the network's weights (Chollet, 2018). The network, layers, loss function, and optimizer are all shown in Figure 2.7.

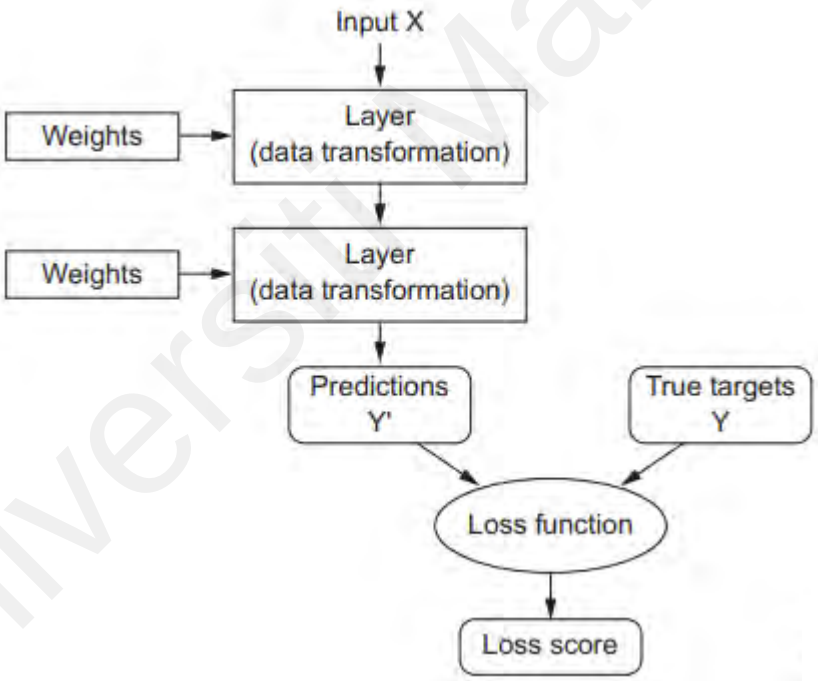


Figure 2.7: The connection between the network, layers, loss function, and optimizer (Chollet, 2018).

Below are the examples of the DL hyperparameters that are normally used and tweaked in order to have an optimum performing DL model.

a) Learning rate

The learning rate, according to Sanderson et al. (2019), is how much the weight parameters are altered at each iteration. One of the most significant hyperparameters is the learning rate. The optimum learning rate is approximately half of the maximum learning rate. One method is to train the model for a few hundred iterations, starting with a very low learning rate (e.g., 10^{-5}) and gradually increasing to a huge value (e.g., 10). For the learning rate, the range of values to evaluate should be less than 1.0 and more than 10^{-6}.

b) Optimizer

Based on the loss function, the optimizer will determine how the network will be modified. By adjusting the weights, it will mould the model into the most exact form possible. Stochastic gradient descent with momentum, Adagrad, Adam, RMSProp, and several other optimizers are examples of optimizers (Chollet, 2018).

c) Batch size

Batch size defines the number of samples that will be passed through the network. It is the hyperparameter of gradient descent that controls the number of training samples to work through before the model's internal parameters are updated (Köse et al., 2020). Higher accuracy of the model is achieved when we use a higher number of epochs, however, it will result in a longer convergence time and overfitting may occur.

d) Dropout

One of the most frequent regularisation approaches for neural networks is to add a drop-out layer during training. It is used to remove units (along with their connections) from the neural network at random to prevent units from over-co-adapting. This lowers overfitting greatly and has significant advantages over other regularisation methods. Dropout is applied to a layer by randomly dropping several output features of the layer during training as shown in Figure 2.8 below. The rate of a fraction of the features that are zeroed out is defined as the dropout rate and is usually set between 0.2 to 0.5 (Srivastava et al., 2014).

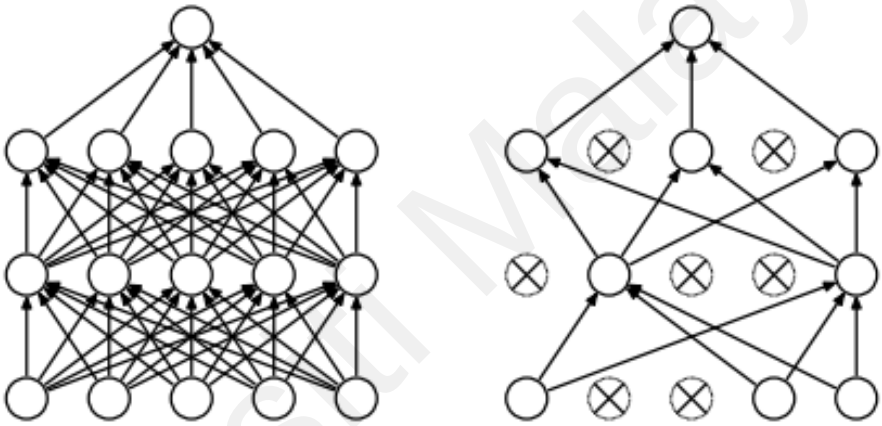


Figure 2.8: Image representation of Dropout Neural Net Model. Left: A standard neural net with 2 hidden layers. Right: Neural Net that applied dropout where the crossed units have been dropped (Srivastava et al., 2014).

e) Epoch

The number of epochs determines how many times the learning algorithm will run over the full training dataset. One or more batches will make up an epoch. When the entire dataset is passed forward and backward through the neural network exactly once, an epoch is completed (Köse et al., 2020). When a larger number of epochs is applied, the model

becomes more accurate; nevertheless, this results in a longer convergence time and the possibility of overfitting.

f) Activation function

According to Chollet (2018), to get access to a much richer hypothesis space that would benefit from deep representations, an activation function that represents the non-linearity is needed.

Without activation function, the dense layer would only consist of two linear operations (a dot product and an addition):

Output = dot (W, input) + b (2.5)

Hence, the layer could only learn the linear transformation of the input data. The hypotheses space of the layer would be the set of all possible linear transformations of the input data into a 16-dimensional space, which in turn restricted the hypothesis space. ReLU activation function is a good default for all hidden layers.

2.5.2.4 Deep Learning Features Engineering and Features Selection

According to Chollet (2018), conventional DL eliminates the necessity for most feature engineering. This is due to the neural networks' capacity to extract valuable features from the raw data presented automatically. Although a DL model may be fitted by validating each weight (*Wk*), the way DL models interpret the variables and the risk score choice is crucial. As a result, it is known as the "black box" (Kwon et al., 2019a).

2.5.2.5 Overfitting and Underfitting

According to Brownlee (2018) overfitting is the model learns the training dataset too well. The model performs well in the training dataset but not on testing samples. When deep network training is adversarial robust, overfitting is a common occurrence, which

means that subsequent training will continue to reduce the classifiers' robust training losses while increasing test losses at a certain point (Rice et al., 2020). The solution to overfitting is to expand the model's capacity. Adding extra nodes or layers to the model is one method of increasing its capacity.

Underfitting was also characterised by Brownlee (2018) as a model that did not learn the problem well enough and performed poorly on both training and holdout samples. The underfitting model was unable to achieve a low enough error rate on the training set. When there are too few neurons in the buried layers, underfitting occurs.

2.5.2.6 Differences between Machine Learning and Deep Learning

DL and ML differ in the way that data is presented in the system. ML almost always requires structured data while DL will rely on the layers of ANN. Besides that, DL is also different from ML in how representations are learnt from the raw data. ML "learn" to act by understanding the labelled data and producing new results with more datasets. DL, on the other hand, places data in a hierarchy of different concepts in multiple layers, which helps them to learn the representations of data with multiple levels of abstraction (Miotto et al., 2018).

In DL, feature extraction can be done automatically while in ML, understanding of features is needed to represent the data. Hence, DL acts similarly to how the human brain works to solve problems. After running the queries through multiple hierarchies of ideas, questions will be related to getting the answers. In ML, however, most of the applied features need to first be identified by an expert or data scientist.

Instead of performing feature selection, DL employs feature learning. Feature learning can only learn all of the characteristics that have been presented and perform the tasks that

have been assigned to it, such as classification and detection, in order to gain significant features that can be used to predict the outcome (Kwon et al., 2019a). On the side of dependency on data, DL will perform well on large datasets while ML perform well on small and medium datasets. Hence, since DL is unable to perform feature selection, the goal of this study is to combine ML feature selection with DL.

2.5.3 Artificial Intelligence Model Development

According to the literature, there are three fundamental steps of ML building methods in health care in general which comprise an exploration phase, solution/design phase, and implementation/evaluation phase (Verma et al., 2021). This study adopted similar techniques or sequences of model development. This method can be implemented in various systems (even though it may slightly differ in the model development phase) because the steps are mostly the same. In order to create and deploy machine-learned solutions in health care, a multidisciplinary relationship between technical specialists and end-users, including physicians, administrators, statisticians, and patients and their families, is required.

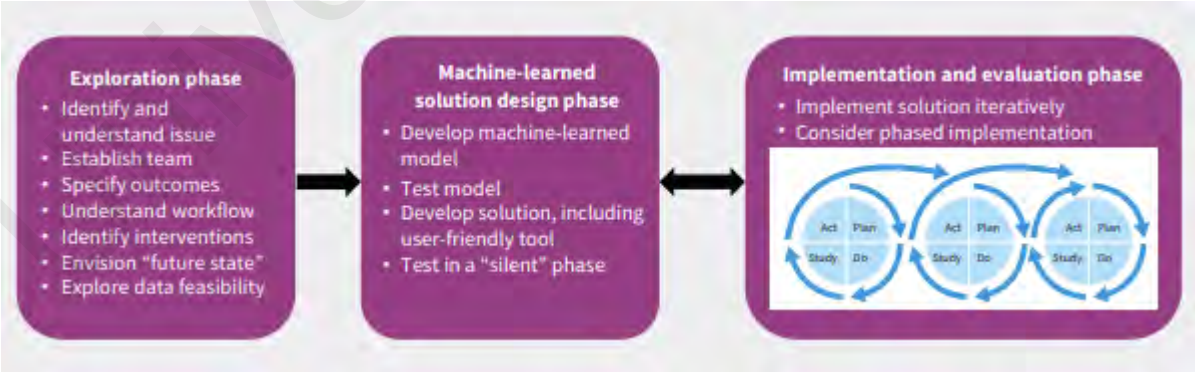


Figure 2.9: Image representation of three basic steps of machine learning model development in health care (Verma et al., 2021).

The same method is specifically parallel in ML model development generally. The processes in using ML algorithms proposed by Goldstein et al. (2017) included data preparation, prediction model execution, model validation, and model testing. The summary of the model development steps for ML can be referred to in Figure 2.9 above.

The outcome must be defined to be predicted as well as the predictor variables that will be used in the model developed during the data preparation process. It is possible to test algorithms and tuning parameters for testing. After that, a loss function can be specified. It is necessary to define rules for imputing and transforming data.

The cross-validation process is set up when running the model prediction to test the tuning parameters and method throughout each iteration. Imputation and standardisation can both be done at the same time. After that, each algorithm's loss will be determined. The cross-validation process was re-run after assessing the variable importance, the categories of cases that were poorly forecasted, and the variable importance.

The testing process included evaluating the performance of the testing set and comparing the metrics to those that were available. Figure 2.10 below summarises the flow of an ML model development.

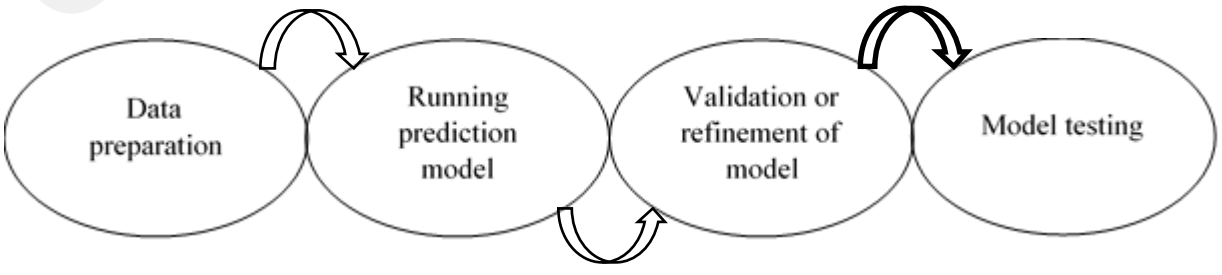


Figure 2.10: Fundamental machine learning model development.

2.5.3.1 Data Preprocessing

a) Zero variance

Columns with only one observation or value are unlikely to be useful for modelling. A single value for all rows in a column does not contain any information for modeling. Variables with a single value can also cause errors or unexpected results which is depending on the choice of data preparation and modeling algorithms. These columns or predictors are known as zero-variance predictors because the variance (average value from the mean) would be zero if we measured it (Brownlee, 2020). Any variables or columns that have a single value should probably be removed from your dataset (Kuhn & Johnson, 2019).

In other situations, the columns could consist of only a few numerical unique values. This situation commonly happens in categorical and ordinal variables because the dataset will only contain numerical values. These columns or predictors are referred to as near-zero variance due to their nature that only consists of a small amount in variation (such as two values for binary dummy variables) and as their variance is not an actual zero but near to the value zero. According to Kuhn and Johnson (2019), near-zero variance columns or predictors tends to have near-zero variance value during the resampling process. These predictors are likely to have small valuable predictive information and are commonly suggested to be removed from the whole dataset. Variables with minimal numerical values can also result in errors or unexpected output.

b) Feature scaling

Different characteristics have different value ranges, according to Bollegala (2017). As a result, feature scaling must be completed prior to the training of any supervised classifiers.

Feature scaling has been demonstrated to boost performance in various classification methods.

Normalization and standardisation are two types of feature scaling processes. Data normalisation is defined by Singh and Singh (2020) as one of the pre-processing steps in which data is scaled or changed so that it contributes evenly to each feature. The fundamental goal of normalisation is to reduce the bias of those features whose numerical contribution to pattern classification is higher.

The use of min-max normalisation is a common way to normalise data. The minimum value of each character is converted to a 0, the highest value is converted to a 1, and the remaining values are turned into a decimal between 0 and 1. The equation of normalisation of a variable value can be referred to as equation 2.6 where, *X*, is the original value, *min(X)*, is the minimum value in the variable, and *max(X)*, is the maximum value in the variable.

$$Xnorm = \frac{X - \min(X)}{\max(X) - \min(X)} \tag{2.6}$$

On the other hand, standardisation, also known as Z-scores normalisation, is a scaling approach in which the values are centred around the mean with a unit of standard deviation, implying the attribute's mean in the resultant distribution becomes zero with a unit of standard deviation. The μ in equation 2.7 is the mean value in the variable and the σ, is the standard deviation of the values in the variable.

$$Xnorm = \frac{X - \mu}{\sigma} \tag{2.7}$$

c) Data balancing

Class imbalance according to Mpanya et al. (2021), refers to the disproportionality of the data classes utilised to train the prediction model, a widespread issue that is not limited to medical data. When the training data for the negative outcome (e.g., dead) has much fewer observations than the majority class (e.g., alive), the classification algorithm is disposed to favour the majority class. This creates complications since the minority class, which bears the weight of the outcome, will have a low accuracy score. Fortunately, class imbalance concerns can be addressed through data manipulation, algorithm manipulation, or a combination of the two (Rekha et al., 2019).

One of the most direct approaches to address class imbalance is to rebalance class distributions. Three fundamental strategies exist for balancing class distributions:

- **Under-sampling**: these techniques are intended to balance the data set by excluding instances of the dominant class.
- **Oversampling**: these techniques replicate instances of the minority group in order to achieve a more equal distribution.
- **Both under- and over-sampling**: This method generates new minority class instances by interpolating between many closely related minority class instances and at the same time reducing the number of instances in the majority class that are considered unimportant.

Under-sampling can result in the obliteration of potentially relevant data, while oversampling increases the chance of overfitting, as the majority of oversampling approaches create perfect copies of minority class samples. Thus, a symbolic classifier, for example, may provide rules that appear to be accurate but cover only one replicated case.

However, for both under- and over-sampling methods, it generates as many minority class instances as necessary to balance the class distributions by allocating 50% of the training examples to the minority class. While this method does not always produce optimal results, it frequently produces results that are comparable to, if not superior to, those produced by using natural class distributions. This method also avoids overfitting and result in the minority class's decision boundaries spreading further into the majority class space. This heuristic employed in this method is intended to mitigate the disadvantages previously discussed.

d) Cross-validation

Cross-validation is one of the most extensively used data resampling strategies to modify model parameters and determine genuine prediction error of models, according to Berrar (2019). Cross-validation is a resampling technique for evaluating ML models on small data sets.

In k-fold cross-validation, the available learning set is first partitioned into k-numbers of subsets with approximately equal size. Fold here refers to the numbers of subsets that are likely to be randomly partitioned. The model will be trained with the k-1 subset which represents the training set. For example, if k= 10 are chosen, the learning set is partitioned into 10 partitions. Then, the dataset will be divided into 10 subsets. Next, one subset, *s1* will be used as the testing set while the other 9 *(s2,...,s10*) sets will be the training set. A new predictive model is built based on 9 training observations and a prediction, *^y1* is made based on the one subset that is used for the testing set. A smaller variance will be attained by using cross-validation rather than a single hold-out of the set evaluator. It also assists in the reduction of variance by more than ten divisions. Overfitting is reduced as a result of this. In Figure 2.11 below, it demonstrates on how 10-fold cross-validation works.

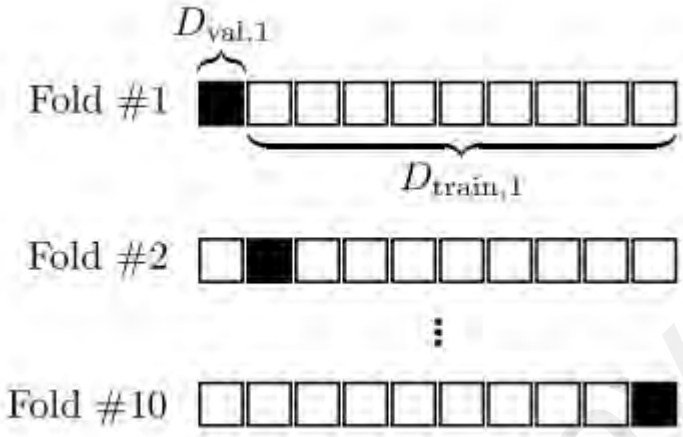


Figure 2.11: Illustration of 10-fold cross-validation where the partitioning of the data is done randomly and D_{train} is the subset for training while D_{val} is the subset for testing (Berrar, 2019).

Model building and hyperparameter tuning operations could also benefit from k-fold cross-validation in the training set. According to Hsieh et al. (2018), they utilised train-test split to split the data and then used k-fold cross-validation with k=10 to do hyperparameter selection for configuration on each model they constructed.

2.5.3.2 Feature engineering

Chollet (2018) defines feature engineering as "a process of implementing hard-coded data changes before data is fed into a model to make algorithms operate better." This is accomplished by combining data knowledge with ML methods.

The feature selection method is only applicable to ML and not to DL, as DL performs feature learning. Due to the neural network's capacity to automate the feature extraction process from raw data, modern DL has helped to reduce the majority of feature engineering

activities. However, when MLP is employed to train a model, feature engineering is considered because problems could be solved more elegantly with fewer resources and data. The DL models' ability to learn features on their own is reliant on the availability of a large amount of training data. As a result, if only a few examples are available for training, the usefulness of their characteristics became critical.

a) Variable importance

When some features are uninformative or heavily associated with other features, the performance of learning algorithms can degrade. Additionally, the ranking of features based on their importance might provide significant information to the end-user. Prior to feature selection, all variables must be ranked from most important (at the top) to least important (at the bottom). The ranking of these variables is essential for the feature selection procedure that follows, as several of the methods used to eliminate variables during the variable reduction process that either will add (forward direction) or delete (backward direction) variables dependent on their ranking (either form the top or the bottom). Given that we used SVM, RF, and LR in this investigation, the next section details on how each algorithm ranked the variables based on their importance.

i. Variable Importance in Support Vector Machine

To pick a suitable subset of SVMs, the kernel-based criterion proposed with the linear combination of features is suggested. As with linear regression, the *i*th coefficient denotes the fraction of total variance around the mean value of the dependent variables that can be explained by the linear relationship between the *i*th variable and the dependent variables. Thus, employing the magnitude of the

coefficient as a variable ranking criterion causes individual variables to be ranked according to their linear fitness. The matching coefficients are then derived based on the highest separability in the feature space. Additionally, the magnitudes of these coefficients are used to assign feature importance. The coefficients are explained by the nonlinear relationship between the class separability of the *i*th characteristic and its class separability (Kuo et al., 2013). The coefficient of separability in the feature space will be determined by the kernel functions (linear, polynomial, or radial).

ii. Variable importance in Random Forest

Breiman (2001) early prototypes of RF software had a variety of options for calculating variable importance. One method for classifying forests involves estimating variable importance based on the forest-averaged decrease in Gini impurity. However, it has dwindled in popularity over time (Grömping, 2009; Louppe et al., 2013). By far the most often utilised measure of importance is a measure termed permutation importance developed by the Breiman-Cutler software. Unlike Gini importance, which assesses significance using in-sample impurity, permutation importance employs a prediction-based technique, utilising the variable's prediction error. Instead of employing cross-validation, which is computationally costly in forests, permutation importance assesses prediction error by utilising out-of-bootstrap cases. Take into consideration that each tree is constructed using a bootstrap sample of the initial data. This data is referred to as OOB data, and the prediction error derived from it is referred to as OOB error (Breiman, 1999). Permutation importance permutes a variable's OOB data and compares the resulting OOB prediction error to the original OOB prediction error—

the rationale being that a large positive value indicates a predictively important variable (Ishwaran & Lu, 2019).

iii. Variable Importance in Logistic Regression

The LR's feature selection algorithm, which performs stepwise regression analysis, employs a different technique for determining the variable's importance without ranking. It is further discussed in detail in the “Wrapper method” part on the feature selection subchapter below.

a) Feature selection

The critical difference between statistical and ML methods is that the former primarily aids in the understanding of relationships between a small number of variables, whereas the latter aids in the identification of new variables from data and improves prediction (Shameer et al., 2018). In ML, the relationships between variables might be difficult to decipher, especially in a “black box” architecture. The time it takes to analyse the data is faster with the use of computational power than with the conventional approach. ML can be used to choose features, classify them, or do both.

Miao and Niu (2016) defined feature selection as a dimensionality reduction strategy. Feature selection is the process of identifying a small subset of key features from a larger set of characteristics to eliminate redundant, irrelevant, or noisy data and improve learning performance. The number of current variables is reduced to a small number, which reduces data dimension, increases efficiency, improves classification precision, enhances information visualisation, mutual exchange of the derived classification models, reduces training time, and improves the accuracy of findings, resulting in more comprehensible and acceptable results (Jović et al., 2015; Kumbhar & Mali, 2016).

The feature reduction method produces variables that help models perform better. The feature selection process can be broken down into four stages: generation (to generate the next candidate subset), evaluation (to evaluate the subset under examination), stopping criterion (to decide when to stop), and validation (to check whether the subset is valid) as seen in the Figure 2.12 below (Liu & Yu, 2005). Filter, wrapper, and embedded are the three general approaches to feature selection (Jović et al., 2015).

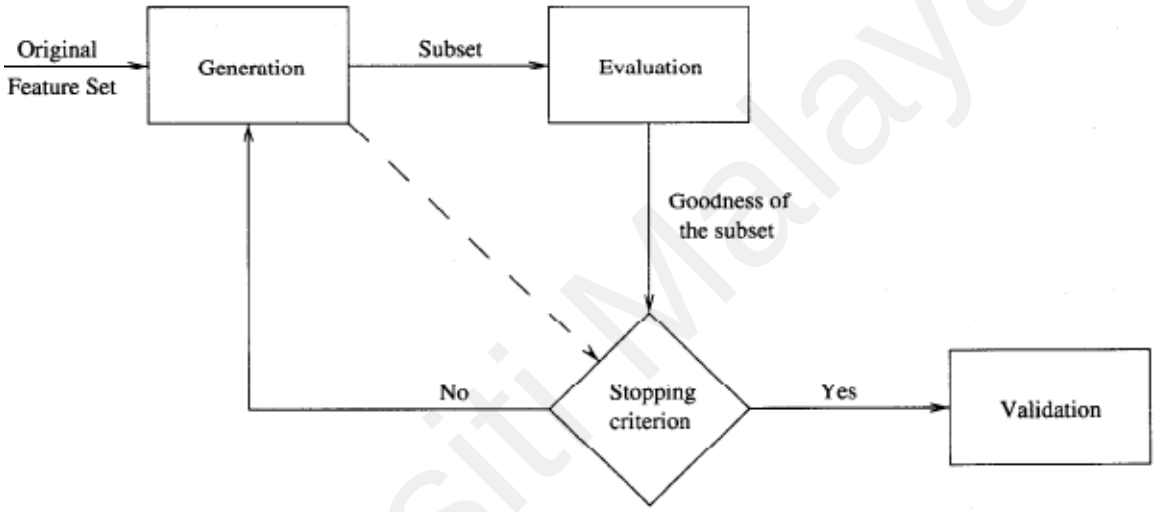


Figure 2.12: Feature selection architecture (Dash & Liu, 1997).

i. Filter Method

The filter technique incorporates an independent measure for evaluating feature subsets without using a learning algorithm (Kumar & Minz, 2014). This method is both efficient and quick to compute (computationally efficient). According to Chandrashekar and Sahin (2014), variable ranking strategies are used as the primary factor for variable selection by ordering in filter methods. Due to their simplicity, ranking algorithms are used, and good

results have been recorded in practical applications. The variables are evaluated using a suitable ranking criterion, and variables that fall below the threshold are removed. Figure 2.13 below summarises the steps of the filter method. Since they are used before classification to filter out the less important variables, ranking methods are filter methods. Some of the examples of filter methods are FOCUS (Almuallim & Dietterich, 1991), ABB (Liul et al., 1998), and relief (Kira & Rendell, 1992).

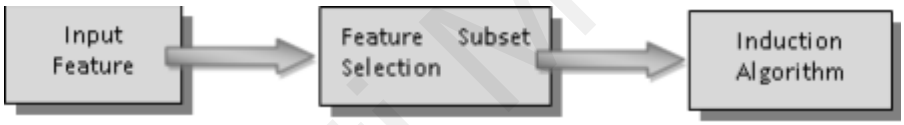


Figure 2.13: The flow of the filter method in feature selection (John et al., 1994).

A fundamental attribute of a unique feature is that it contains meaningful information about the data's many classifications. This property is known as feature importance, and it is a measure of a feature's utility in distinguishing between various classes (Kohavi & John, 1997). When a feature can be independent of the input data but not of the class labels, it is regarded as significant; nevertheless, a feature that does not influence the class labels can be eliminated (Law et al., 2004).

As previously stated, inter-feature correlation is critical in determining unique characteristics. The underlying distribution is unknown in real applications; therefore, the classifier accuracy is used to assess it. As a result, an ideal feature subset may not be unique, as the same classifier accuracy might be achieved with several combinations of features.

Saeys et al. (2007) on the other hand, pointed out that filter methods have a poor lengthy interaction with classifier algorithms, and they also stated in their study that because most filter methods are univariate in nature, these approaches may not pay attention to the values of other variables. The disadvantages of filter methods include redundancy of the selected features and the failure to evaluate some crucial correlations between features, as features can obtain a poor score when employed by the ranking algorithm. Filter methods also can miss features that are not valuable on their own but are extremely useful when paired with others (Chandrashekar & Sahin, 2014).

ii. Wrapper Method

Evaluation criteria are the only way to distinguish the filter and wrapper approaches. For subset evaluation, the wrapper technique employs a learning algorithm (Kumar & Minz, 2014). The contribution of each feature to the performance of a certain type of classifier is taken into account during feature selection. The wrapper methods class uses a set of defined criteria to make a step-by-step selection of variables using the forward/backward criterion. Despite their slowness, wrapper methods are suitable for final model construction when compared to other methods. Sequential Forward Selection, Sequential Backward Selection (SBE) (Koller & Sahami, 1996), and Recursive feature elimination (RFE) (Kohavi & John, 1997) are all common wrapper approaches.

Forward Selection is an iterative approach in which a model starts with no features or an empty set, and then features are added to the model in each iteration, improving the model's performance until all of the variables are included in the model, or until adding features does not affect the model's performance. The predictor that improves the model the most is added at each iteration (James et al., 2013; Kabir et al., 2010). SBE is a sequential forward selection alternative. To increase model performance, this SBE starts with a full set of variables and removes the least significant variable at each iteration. The technique is repeated until there is no more improvement in the elimination of characteristics (James et al., 2013). Because weaker features are not considered during subset selection, forward selection finds a weaker subset of features. Furthermore, forward feature selection has lower computational complexity than backward feature selection. Forward feature selection method errors made early in the process are not repaired later (Kumar & Minz, 2014).

Recursive Feature Elimination (RFE) (Guyon et al., 2002) is a wrapper method that aims at finding a minimal and best-performing set of variables, which leads to a good prediction model. RFE will train the classifier and compute the ranking criteria for all features while removing the feature with the smallest ranking criterion or not important. The features that are top ranked (eliminated last) are not necessarily the ones that are individually most relevant. It repeatedly creates a model and keeps aside the best or the worst performing features at each iteration. It then ranks the features based on the order of their elimination. It then constructs the next model with the remaining features until all the features are exhausted. This iterative procedure is an example of backward feature elimination (Kohavi & John, 1997). It should be noted that RFE has no consequence on correlation methods since the ranking measure is calculated with information about a single feature.

In R, a set of functions must be specified using *rfeControl$functions* for each model. There are pre-defined sets of functions for a variety of models, including linear regression (*lmFuncs*), random forests (*rfFuncs*), naive Bayes (*nbFuncs*), bagged trees (*treebagFuncs*), and functions that may be used with caret's train function (*caretFuncs*). The latter is beneficial if the model involves tuning parameters that must be determined at each iteration (such as SVM). The most commonly used function for RFE in the ML model training is the "random forest" (*rfFuncs*) because it has a nice built-in mechanism for computing feature importance (Kuhn, 2009). Several studies in the medical fields, including Aziida et al. (2021), Das et al. (2020), Bahl et al. (2019), and Macias et al. (2020), have been conducted to demonstrate the robustness of the *rfFuncs* in feature reduction for the benefits of reducing the cost of ML model development and, as a result, improving the ML model performances. The *rfFuncs* is further discussed in the Methodology section of the thesis.

Stepwise regression is a technique for fitting regression models in which the selection of predictive variables is automated (Efroymson, 1960). Each stage considers whether a variable should be added to or subtracted from the set of explanatory variables based on some predetermined criterion (Draper & Smith, 1998). The main approaches for stepwise regression are forward selection, backward elimination, and bidirectional elimination. Starting with no variables in the model, forward selection involves testing the addition of each variable using a chosen model fit criterion, adding the variable (if any) whose inclusion gives the most improvement in the fit, and repeating this process until none improves the model significantly. Backward elimination entails starting with all candidate variables, testing their deletion using a chosen model fit criterion, deleting the variable (if any) whose loss results in the least deterioration of the model fit, and repeating this process until no more variables can be deleted without an insignificant loss of fit. Bidirectional

elimination, a combination of the foregoing, with factors to be included or removed tested at each phase.

Instead of using the traditional way of stepwise regression analysis using t-tests or F-tests fitting the final selected model based on the p-value, StepAIC is one of the most extensively used feature selection for stepwise regression analysis method in R. The Akaike information criterion (AIC) is a predictor of prediction error and, as a result, of statistical model quality for a given set of data (Aho et al., 2014). AIC measures the quality of each model in relation to the other models given a set of data models. As a result, AIC can be used to choose a model. The information theory supports AIC. When a statistical model is used to describe the process that created the data, it is virtually never accurate; thus, some information is lost when the model is used to represent the process. The AIC calculates the amount of information lost by a given model: the less information lost, the greater the model's quality. AIC considers the trade-off between model goodness of fit and model simplicity when assessing the amount of information lost by a model. To put it another way, AIC considers both the risks of overfitting and underfitting. The AIC value is calculated as follows;

$$AIC = 2k - 2\ln(L) \tag{2.8}$$

where k is the number of the estimated parameters in the model and L is the maximum value of the likelihood function for the model (Akaike, 1985).

The AIC value is examined to see if it is increasing or decreasing as more variables are added or discarded. To arrive at the final set of features, the StepAIC value is aimed to be as low as possible. Hence, we can say that AIC provides a means for model selection. As a result, while AIC rewards goodness of fit (as measured by the likelihood function), it also

contains a penalty that grows in proportion to the number of estimated parameters. Overfitting is discouraged by the penalty, which is desirable because increasing the number of parameters in a model almost always enhances the fit's goodness.

The wrapper method chooses the most appropriate subset for the learning algorithm. As a result, the wrapper technique usually performs better (Kumar & Minz, 2014) than the filter method. Figure 2.14 below shows the summary of how wrapper method is performed.

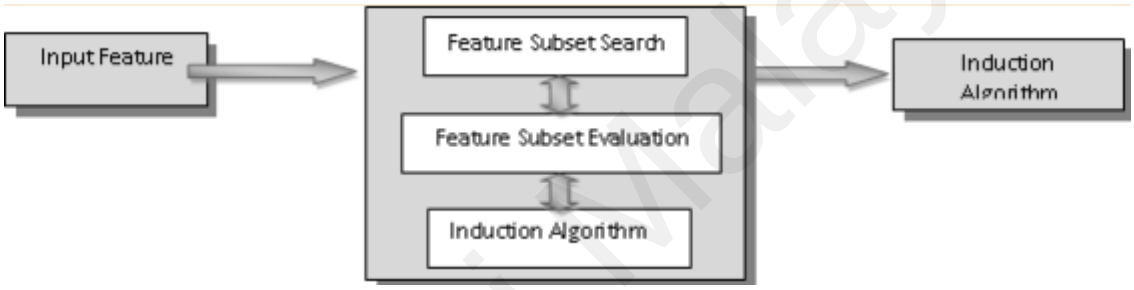


Figure 2.14: The flow of the wrapper method in feature selection (John et al., 1994).

iii. Embedded Method

The embedded approach has a lower computing cost than the wrapper approach when interacting with the learning algorithm. It also records the interdependencies between features (Blum & Langley, 1997). The most common strategy is to include feature selection in the training process. It considers not just the correlations between input and output attributes, but also looks for features that might help with local discrimination. The best subsets for a known cardinality are determined using independent criteria (John et al.,

1994). The learning process is then used to choose the final optimal subset from among the optimal subsets of various cardinalities.

2.5.3.3 Data imputation

The majority of statistical and ML techniques are insufficiently robust when dealing with missing variables. Missing data has an impact on them. Missing data provides an element of ambiguity into data analysis, which can impact statistical estimator qualities, resulting in a loss of power and erroneous findings (Schmitt et al., 2015; Somasundaram & Nedunchezhian, 2011). Although a number of variables impact the quality of an ML algorithm's output, such as feature selection, algorithm selection, sampling techniques, training, test, and validation datasets, one of data scientists' main concerns is how to cope with missing data (Brown & Kros, 2003).

Dealing with missing values effectively is a difficult undertaking that necessitates which are a thorough review of all instances of data to detect patterns of missingness in the data and a thorough understanding of various imputation strategies.

Data that is missing could be due to equipment failure, data that is conflicting with other data and thus destroyed, data that is not recorded due to misunderstanding, or data that is not judged important at the time of data collection.

Before implementing any strategy for dealing with missing data, it is vital to understand why it's missing. Missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) are three possible missing data techniques proposed (Little & Rubin, 2002; Rubin, 1976).

- **MCAR:** It is the highest level of randomization, suggesting that the pattern of missing values is fully random and unrelated to any variable in the study. As a result, if missingness is unrelated to any data in the dataset, it means data is missing at random. The MCAR assumption states that the likelihood of missingness is independent of both observed and unobserved values of any variable in the dataset.
- **MAR**: In this case, the observed data in the dataset determines the likelihood of missing data. It indicates that observable data determines the likelihood of missingness, while unobservable data does not. The missing value of any variable in the dataset is reliant on the observed values of other variables in the dataset because there is a relationship between the attribute carrying the missing value and other attributes in the dataset. The observed values in the dataset may reveal the pattern of missing data.
- **MNAR**: Missingness is determined by unobserved data rather than observable data in this scenario. Consequently, missingness is defined by either missing data or the item itself, due to the response variable's high sensitivity to answer. When using MNAR data, the likelihood of missing data is proportional to the value of the missing data. The missing data pattern is not random and is unpredictably different from the observed values of the other variables in the dataset.

There are two approaches to dealing with missing data (Han et al., 2011). The first technique is to simply ignore missing data, while the second option is to think about imputation.

a) **Ignoring missing data**: Missing data ignoring method is a strategy that simply ignores occurrences when data is missing. They are frequently utilised, and they are often the go-to solution for dealing with missing data. This method has a serious problem in that it does not use the entire dataset. This is a suitable choice when the dataset includes a small number of missing values. There are two techniques to ignore missing data in general:

- Listwise deletion is the method used for the full case analysis. All observations with missing values for any variable of interest are removed in a comprehensive case analysis. As a result, this method limits the analysis to observations with all values recorded, which typically leads to a skewed estimate and a loss of precision (Schafer & Graham, 2002).
- A pairwise deletion analysis is one in which all situations in which the variables of interest are present are examined. It does not omit a whole unit, but it does make use of as much data as feasible from each unit. Even if any of the variables have missing values, this technique has the benefit of saving as much data as possible for analysis. The downside of this strategy is that it uses different sample sizes for different variables (Schafer & Graham, 2002). Each individual analysis has a larger sample size than the entire case study.

b) **Imputation of missing values**: Imputation of missing values is a process for replacing missing values with plausible alternatives (Rubin, 1976). The numerous imputation strategies strive to deliver accurate population parameter estimation without reducing the capacity of data mining and data analysis techniques. The

quantity of data missing determines the appropriate approach for missing data. Although there is no hard and fast rule about how much missing data is bad, it is usually a good idea to compare the findings before and after imputation if more than 25% of data is missing. As a result, the imputation method's performance is unaffected by the dataset or the percentage of missing values (Jadhav et al., 2019). According to Cismondi et al. 2013, deletion of missing values is a typical procedure in the medical industry. In some datasets, the percentage of missing data can reach 50% or higher, and in these cases, imputed data seems to be inherently wrong.

The data scientists' primary focus is data quality. Although data quality is influenced by some factors, one of the most important is data incompleteness. As a result, data scientists must deal with missing data concerns with rigour before analysing data and allowing end-users of data mining initiatives to make viable conclusions. Data imputation is a technique for replacing missing values with the most plausible values to make data complete and suitable for analysis by replacing missing values with the most plausible values.

There are two types of data imputation methods: single imputation methods and multiple imputation methods.

i) Single Imputation*:* A single missing value is replaced by a unique value retrieved from the entire data is a part using this procedure (Jerez et al., 2010). Some of the examples of single imputation methods are:

- Mean imputation: The mean value of each non-missing variable is used to fill in missing values for all observations in the general approach to mean imputation, which may be thought of as a simple application of regression imputation. The disadvantage of this method is that if there are a lot of missing values, all of them are replaced with the same imputation value, which is the mean, which changes the shape of the distribution. When you compare the standard deviation before and after imputation, it gets smaller. The median and modus imputation procedures are similar to mean imputation. These techniques were developed to account for the imputation of data that was not normally distributed.

- Regression imputation: This is a more advanced version of the single imputation technique. Missing values are replaced with predicted data using regression based on non-missing data from other variables in this method. This strategy is based on the premise that the attributes have a linear connection. However, because most relationships are not linear, applying regression to replace missing values would bias the model. This approach has the potential to yield skewed findings, especially when using MNAR and MAR (Schafer & Graham, 2002).

- kNN imputation: Missing values are imputed using this method by copying values from similar entries in the same dataset. A distance function is used to determine how similar the two qualities are. It is not necessary to create a prediction

model for each attribute, but it does have drawbacks. Because of the k value selection, evaluating huge datasets takes a long time.

- Hot deck imputation: For the imputation of an incomplete instance, hot-deck imputation techniques use a completely observed donor case. The donor case's related values are used to fill in the gaps. The various hot-deck procedures are defined by the various methods of locating a donor case (Molnar et al., 2008).

ii) Multiple imputations: In a nutshell, the objective of this method is to replace each missing value, *m* with $m > 1$ plausible value. The *m* complete data sets are then evaluated independently using typical full data processes, and the *m* sets of results are then integrated using Rubin (1988) formula to generate a single overall set of results. The multiple imputation method anticipates missing data using all of the data set's available information. Since variables in the data set are frequently connected with insufficient variables or associated with missing value, multiple imputations is often better equipped to provide sufficient parameter estimates than ad hoc methods such as unconditional mean imputation, which cannot take advantage of this information (Schafer & Graham, 2002). The mean of the *m* estimations is the multiple imputation point estimate. The variance estimate is calculated by combining within-imputation variability (the mean of the missing values standard error estimates) and between-imputation variability (the standard deviation of the missing values point estimates). The between-imputation variance component directly represents the uncertainty about the parameter estimate owing to unobserved information, and as the missing

value estimates go closer together, variance becomes bigger, resulting in a broader confidence range, and therefore bias about the parameter estimate may be mitigated.

- Multiple Imputation by Chained Equations (MICE) is an increasingly popular method for doing multiple imputations (Royston & White, 2011). Predictive Mean Matching (PMM) is a common approach for doing multiple imputations on continuous or semi-continuous multivariate (Van Buuren, 2011; Vink et al., 2014). This study adopted PMM because according to Morris et al. (2014), the algorithm iteratively imputes each missing column (target) by producing synthetic values from the dataset's other variables (predictors). Missing value predictors are filled with an initial value before being completed with the most recently generated imputations. There is a different imputation mode for each column. By imputing an actual observed value, PMM is frequently superior to fully parametric multiple imputation approaches in preserving the fundamental data distribution and the relations within the set of data, and therefore better able to screen any bias caused by failed distributional assumptions.

According to a few studies of the literature, the performance of the proposed imputation approaches is highly dependent on the issue domain (e.g., number of instances, number of variables, missingness patterns), and there is no clear indication that one method is superior to the others (Perez et al., 2002; Taylor et al., 2017). In the design of a scoring system for predicting death in ICU patients, Perez et al. (2002) introduced single, hot-deck, and multiple imputation approaches to impute missing data. Differences in areas under the ROC curve were statistically significant but not clinically important, according to the findings. Taylor et al. (2017) investigated the effects of seven alternative imputation approaches on multiple biological matrix analyses (half minimum, mean, kNN, local least squares

regression, Bayesian principal components analysis, singular value decomposition, and RF) and they concluded that no imputation approach is better than the others, but the mean and half minimum performs poorly. Table 2.11 summarises the advantages and disadvantages of single and multiple imputation (Libasin et al., 2020).

Table 2.11: The advantages and disadvantages of single and multiple imputations.

Method	Advantage	Disadvantage
Singke Imputation	• Simple to apply and understand • Quick • Applicable to any statistical analysis • There are no specific computational methods that must be used. • Many statistical software programmes use it by default. • Under the MCAR assumption, unbiased parameter estimations	• A significant reduction in the sample size • Statistical power is reduced. • Doesn't make use of all of the facts • If the data is MCAR, there is a command loss. • If the data isn't MCAR, the findings will be skewed. • The relationship between variables can be influenced.
Multiple Imputation	• Approach in general • Easy to comprehend, but difficult to programme • Ad hoc techniques lacked the legitimacy of unbiased estimates. • Ensures that sample size and statistical power are not harmed. • There is software available for purchase. Estimates' standard errors can be estimated.	• Programming is difficult (specific software is required). • Extensive computing • The imputation model is difficult to regulate for the analyst. • The variance is higher than it is for single-imputation approaches.

2.5.4 Performance Evaluation

The commonly employed metrics in evaluating the performance of a diagnosis model, according to Kim et al. (2017), comprised accuracy, sensitivity, specificity, likelihood ratio, and area under the ROC curve (AUC) (Kim et al., 2017). The following metrics are used in this study to evaluate model performances.

2.5.4.1 Confusion Matrix

A confusion matrix is used to evaluate the performance of classifiers on datasets (Hasnain et al., 2020). The confusion matrix is commonly used in ML for supervised classification or determining the behaviour of classification models (James et al., 2013). A confusion matrix's square structure is represented by rows and columns, according to Caelen (2017). 2 × 2 matrices are used to represent a confusion matrix in binary classification. In a confusion matrix, four metrics were used: 'true positive' (TP), 'true negative' (TN), 'false positive' (FP), and 'false negative' (FN). A confusion matrix with the *k* class has a *k* x *k* confusion matrix in the multiclass issue.

True Positives (TP) show that the patient has a disease with a positive prediction, but False Positives (FP) indicate that the patient does not have a condition but does have a positive prognosis. True Negative (TN) means the patient does not have an illness and has a negative prediction, whereas False Negative (FN) means the patient does have a condition but has a negative prognosis.

The accuracy of a diagnosis model refers to its capacity to correctly identify patients with and without the disease. Sensitivity refers to the model's capacity to correctly identify patients with the disease, while specificity refers to the model's ability to accurately identify patients without the disease.

	Actual (Yes)	Actual (No)
Prediction (Yes)	True positive (TP)	False-positive (FP)
Prediction (No)	False-negative (FN)	True negative (TN)

Figure 2.15: The confusion matrix.

The diagonal items (TN, TP) in Figure 2.15 are true predictions, whereas the others (FP, FN) are false predictions (Catal, 2012). Other performance evaluation metrics will be calculated using these values. Wallert et al. (2017) found that when using ML to predict two-year survival following a first myocardial infarction, false negatives (non-survived patients anticipated to survive) were more expensive than false positives (survived patients predicted to die) for mortality prediction.

a) Accuracy

One of the most popular approaches to assess a model's performance is to look at its accuracy. Saura (2021) described accuracy as a model's or method's quality and accuracy.

The ratio of all correct predictions made to the total number of forecasts made can alternatively be described as accuracy (Jain & Singh, 2018). AUC greater than 0.70 shows that the predictive model proposed a good discriminatory ability, whereas AUC less than 0.50 suggests that the predictive model proposed a low discriminatory ability (Mpanya et al., 2021).

The formula for calculating accuracy is as stated in equation 2.9 below where TP, is true positive, TN, is true negative, FP, is false positive and FN is false negative.

$$Accuracy = \frac{(\,TP = TN\,)}{(TP + FP + TN + FN\,)} \tag{2.9}$$

b) Sensitivity (Recall or True Positive Rate)

Jain and Singh (2018) defined sensitivity as the ratio of true positive to the sum of true positive and false negative. In medical diagnosis, sensitivity is the ability of a test to identify those with disease correctly. High sensitivity with negative test results indicated that a person may not have the disease.

According to Veropoulos et al. (1999), sensitivity is used to calculate the misclassifications in the positive cases. The formula for sensitivity is shown below.

$$Sensitivity = \frac{TP}{TP + FN} \tag{2.10}$$

c) Specificity (True Negative Rate)

Specificity is defined by Jain and Singh (2018) as the ratio of true negatives to the sum of true negatives and false positives. Specificity is utilised to determine the misclassifications in the negative instances (Veropoulos et al., 1999). The formula for specificity is as follows:

$$Specificity = \frac{TN}{FP + TN} \tag{2.11}$$

d) Predictive values

When test findings are provided to clinicians, the two other indices that are helpful in clinical practise are the positive predicted values (PPV) and the negative predicted values (NPV). PPV is also known as precision. The likelihood of disease for positive test results is known as the PPV, whereas the probability of health for negative test results is known as the NPV. The preceding prevalence of disease in the population has an impact on these two indicators, despite the fact that they are valuable for clinical decision-making. A higher prevalence of the disease results in an increased PPV, however a higher prevalence also results in a decreased NPV (Hajian-Tilaki, 2013). The equations of PPV and NPV are shown in the equations 2.12 nd 2.13 below.

$$PPV = \frac{TP}{TP + FP} \tag{2.12}$$

$$NPV = \frac{TN}{TN + FN} \tag{2.13}$$

e) F-Score

Accuracy in classification is a commonly used metric because it can be reduced to a single, straightforward indicator of overall model performance.The F-Measure presents a method for combining precision and recall into a single measure that is capable of capturing both of these features (Taha & Hanbury, 2015).

After the scores for precision and recall have been determined for a binary or multiclass classification problem, it is possible to combine the two scores in order to determine the F-measure for the problem. When taken by themselves, neither precision nor recall reflect the complete story. It is possible to even have horrible precision but fantastic recall, or it's also possible for us to have excellent recall but terrible precision. The F-measure provides a method for expressing both of these issues through the use of a single score.

The equation to calculate F-score is as mentioned in equation 2.14 below.

$$F - score = \frac{2.Precision.Recall}{Precision + Recall} \tag{2.14}$$

2.5.4.2 Receiver Operating Characteristics (ROC) curve

The use of ROC graphs in the ML field has risen steadily, partly due to the recognition that basic classification accuracy is typically a poor metric for evaluating performance (Faizal et al., 2021). The ROC curve was an effective approach to analyse the effectiveness of the diagnosis test (Kumar & Indrayan, 2011). When the diagnostic test is on an ordinal scale with a minimum of 5 categories or is continuous, the plot depicted the trade-off between sensitivity and (1-specificity) across a succession of cut-off positions. The ROC

curve had the advantage of allowing simultaneous display and comparison of one or more diagnostic tests in one picture.

Due to various reasons, such as imbalanced data, too few variables, changing continuous variables into categorical data (degree of discrete creating bias), and false positive and false negative concerns, accuracy has been determined not to be a good measure of performance in several studies. As a result, measuring and comparing the performance of these models with the area under the receiver operating curve (AUC) is more accurate (Faizal et al., 2021).

To assess the performance of a certain diagnosis test, the AUC was used since it was a combined measure of sensitivity and specificity. With the maximum AUC of 1, it indicated that the diagnostic test was perfect in classifying into diseased and non-diseased. The higher the AUC, the better the performance of the model at distinguishing between the positive and negative classes.

In the study done by Wallert et al. (2017), AUC was taken as the performance metric for their models developed as classes were heavily unbalanced, as it is not imbalance-sensitive (Wallert et al., 2017). Between 0 and 1, 0.5 denotes random guessing, and any feature or variable with an AUC more than 0.7 may be a potentially helpful clinical classifier. In mortality prediction, false negatives are considered more disastrous than false positives. However, the judgement was also made along with the consideration of base rate incidence, consequences of false negatives/positives, test risk, cost, etc. AUC values between 0.7 and 0.8 are considered acceptable, values between 0.8 and 0.9 are considered excellent and values greater than 0.9 are considered outstanding (Mandrekar, 2010).

2.5.4.3 McNemar's test

According to Sun and Yang (2008), McNemar's test, named after Q.McNemar, was introduced in 1947. It is a non-parametric method used on nominal data to test the equality of row and column marginal frequencies.

McNemar's test can also be used to determine if there are any differences between two groups on a dichotomous dependent variable. According to Omolala & Wilella (2012), McNemar's test can be defined as a type of chi-square test that uses dependent (paired or correlated) data instead of independent data (Adedokun & Burgess, 2012).

2.5.4.4 Net Reclassification Improvement Index (NRI)

The AUC has evolved as a result of applications in diagnostic testing in radiology back in the year 1982 (Hanley & McNeil, 1982). The study stated that the area under the sensitivity vs. 'one minus specificity' plot for all potential cut-off values is known as the AUC. This concept has been proven to be identical to defining AUC as the chance that a given diagnostic test (predictive model) assigns a higher probability of an event to those who actually experience (or develop) events. The difference in AUCs is determined using a model with and without the variable of interest is simply defined as the improvement in AUC for a model incorporating a new extra variable. However, the magnitude of this increase is frequently insignificant. For example, a study by Shouval et al. (2017) showed that the ML methods with new variables added to it predicting 30-day mortality after STEMI as compared to GRACE score increase the model AUC from 0.87 to 0.91.

As a result of the foregoing, some researchers began to investigate various ways for assessing the improvement. In the medical literature, reclassification tables are becoming increasingly prominent especially in cardiovascular epidemiology (Chattopadhyay et al., 2018; Kwon et al., 2019a; Mahler et al., 2013; Myers et al., 2017; Wang et al., 2020;

Widera et al., 2013). For example, Myers et al. (2017) compared a model developed to improve risk stratification after ACS using the conventional risk prediction score, TIMI risk score (‘old’ model) with a model that was developed using ANN ('new' model), and the AUC values are 0.670 and 0.743 respectively. The two-category NRI with respect to the TIMI risk score was 0.065 (6.5%) when they classified the predicted risks obtained using their two models (old and new) into three categories and then cross-tabulated these two classifications; over the 1000 bootstrap trials, an average of 87 patients was correctly reclassified using the ANN model with a standard deviation of 9.4 patient (Myers et al., 2017). This is similar to the study by Kwon et al., (2019a) that used the DL method in building a new prediction model for in-hospital mortality and then compare it to the GRACE score, their AUC values are 0.905 and 0.851 respectively. The DL method predicted 34 in-hospital mortality patients more accurately than the GRACE score with the net reclassification of 0.6% (Kwon et al., 2019a).

In 2008, Pencina et al. (2008) established the net reclassification improvement index (NRI), a novel measure of incremental value. The overall NRI is provided as a statistic, which is defined as the sum of the net proportion of people with and without the event of interest who were correctly assigned a different predicted risk. The event NRI (NRI^{e}) is defined as the net proportion of people who properly assigned a higher predicted risk to the event of interest, while the non-event NRI (NRI^{ne}) is defined as the net percentage of people who correctly assigned a lower predicted risk to the event of interest. A simple asymptotic test can be used to determine the significance of the improvement.

According to Pencina et al. (2008), the equation for event NRI is:

$$NRI_e = P(up|event) - P(down|event) \quad (2.15)$$

While the equation for non-event NRI is:

$$NRI_{ne} = P(down|nonevent) - P(up|nonevent) \quad (2.16)$$

And the equation of NRI, in general, is the summation of the two previous equations:

$$NRI = P(up|event) - P(down|event) + P(down|nonevent) - P(up|nonevent) \quad (2.17)$$

Any ‘upward' movement in categories for event subjects (those who are involved in the event) indicates better classification, while any ‘downward' movement indicates worse reclassification. For those that do not experience events, the interpretation is the polar opposite (Pencina et al., 2011). Alternatively, the NRI can be computed by calculating the difference between the proportions of individuals moving up and down for those who develop events, as well as the equivalent difference in proportions for those who do not develop events and then subtracting these two differences. By way of explanation, “up” means that the new risk model assigns a person to a greater risk group than the previous one. Similarly, “down” denotes that a person has been assigned to a lower risk category by

the new model. The risk prediction model with established predictors is known as the "old" model in the definition of " net reclassification indices." The "new" model is one that adds a new predictor.

There are two types of NRI index which are categorical and category free reclassification index (Kerr et al., 2014). For example, $NRI^{0.2}$ is a two-category index with a 0.20 cut-off separating low and high risk. The $NRI^{0.1,0.2}$ is a three-category index with cut-offs of 0.10 and 0.20, indicating low, medium, and high risk. A net reclassification index can be defined using any set of risk thresholds. The “category-free net reclassification index” (also known as the “continuous net reclassification index”) translates the overall definition of NRI to any upward or downward change in expected risks. The category-free index is also denoted by $NRI^{>0}$.

The net reclassification index is frequently misinterpreted as a proportion. For example, interpreting the index as “the proportion of patients reclassified to a more acceptable risk category” is inaccurate because it is *P* (up and event) + *P* (down and nonevent) (Pickering & Endre, 2012). Because NRI^{e} and NRI^{ne} represent proportional differences, they are easier to interpret than the net reclassification index. The net proportion of events assigned a higher risk or risk category is referred to as NRI^{e}. The net proportion of nonevents assigned a lower risk or risk category is known as NRI^{ne}. The word "net" is key in this case for proper interpretation.

2.6 Machine Learning and Deep Learning as Alternative Mortality Prediction Methods

Advances in computer science, as well as the need for precision medicine, have resulted in the accumulation of multidimensional data from several fields (Shameer et al., 2018).

This technique has been used for several tasks on a variety of data formats, including electronic health records (Jensen et al., 2012), pathologic specimens (Komura & Ishikawa, 2018), and gene expression microarrays (Pirooznia et al., 2008), as well as for novel applications like medical robotics (Kassahun et al., 2016). In all the medical data analysis scenarios where ML has been used, it has shown promise as a potentially strong tool for detecting data trends and patterns that would otherwise go unreported if traditional statistical approaches were used.

On the other hand, the medical field has also been adopting ML in its practices to assist in the diagnosis and prognosis of patients. The development of a predictive model using ML is widely used in various medical domains such as diagnosing patients with diabetes mellitus (Juneja et al., 2021), predicting prolonged length of stay in newborns (Thompson et al., 2018), predicting the length of stay of patients in the Intensive Care Unit (ICU) (Picone et al., 2021), early detection of autism (Abbas et al., 2017), diagnosis of arrhythmia using images from electrocardiogram (ECG) (Desai et al., 2015) and recently in Coronavirus disease (COVID-19) cases analysis (Kwekha-Rashid et al., 2021). These predictive models have shown positive feedback in predicting the desired outcomes and resulting in better performance.

The ML methods used in the aid of building a predictive model in the various medical domain are summarised in Table 2.12 below. By analyzing the research done in different medical domains, it can be concluded that different ML suited different medical datasets when performing predictions. For example, in ML-Based Prediction of Prolonged Length of Stay in Newborns by Thompson et al. (2018), RF outperformed other methods with the area under the ROC curve of 0.88.

Table 2.12: Summary of previous studies that incorporate ML in the medical domain.

Paper title and authors	Data sources	Number of instances	Input variables	Model performances metrics
Predicting Diabetes Mellitus with Machine Learning Techniques (Zou et al., 2018)	Hospital physical examination data in Luzhou, China	82,694	14 variables	Accuracy (all features,blood glucose, PCA,mRMR, without blood glucose, 11 features): **Random forest** (0.81,0.76,0.74,0.75,0.72,0.71) **J48** (0.76,0.76,0.74,0.76,0.69,0.69) **Neural network** (0.0.78,0.76,0.74,0.76,0.70,0.70)
Machine Learning-based Prediction of Prolonged Length of Stay in Newborns (Thompson et al., 2018)	Healthcare Cost and Utilization Project (HCUP) dataset.	17,889	20 variables	ROC score: **ZeroR** (0.5) **Naïve Bayes** (0.70) **Logistic** (0.72) **Multi-layer perceptron** (0.72) **Simple Logistic** (0.72) **SVM** (0.58) **J48** (0.78) **Random Forest** (0.88) **Random tree** (0.67)
Predicting length of stay using regression and Machine Learning models in Intensive Care Unit: a pilot study (Picone et al., 2021)	University Hospital of Naples "Federico II" adult and neonatal ICU dataset.	415	5 variables	Mean absolute error: **DT** (0.70) **RF** (0.53) **GBT** (0.56) **SVM** (0.56) **KNN** (0.62) **MLP** (0.70)

Table 2.12, continued.

Paper title and authors	Data sources	Number of instances	Input variables	Model performances
Machine Intelligent Diagnosis of ECG for Arrhythmia Classification Using DWT, ICA and SVM Techniques (Desai et al., 2015)	PhysioNet, *MIT– BIH arrhythmia database*	110,093	5 variables	SVM kernels (F-Score): **Linear** (59.67) **Quadratic** (76.54) **Polynomial** (64.43) **RBF** (68.36)
Machine learning-based approaches for detecting COVID-19 using clinical text data (Khanday et al., 2020)	Open-source data repository GitHub	212	24 variables	F Score: **Logistic regression** (0.95) **Multinomial Naïve Bayesian** (0.95) **Support Vector Machine** (0.86) **Decision tree** (0.92) **Bagging** (0.92) **Adaboost** (0.88) **Random forest** (0.93) **Stochastic gradient boosting** (0.93)

Additionally, AI-based techniques especially ML are being utilised to aid in the development of standardised predictive models that could help cardiologists with patient-specific guidelines and decision-making. These would, in turn, assist clinical providers in reclaiming time and improving patient-provider relationships (Shameer et al., 2018).

Based on the literature, research regarding mortality prediction in ACS was widely conducted in countries around the globe for the past decade adopting ML as the predictive method. Some of the western countries that used this method are the United States of America (USA) (Al'Aref et al., 2019; Barrett et al., 2019; Frizzell et al., 2017; Mansoor et al., 2017; Myers et al., 2017; VanHouten et al., 2014), Sweden (Wallert et al., 2017), and Poland (Pieszko et al., 2019). Countries in Asia are not left out in the advancement of ML as Israel (Shouval et al., 2017), China, Korea (Li et al., 2017; Li et al., 2020), and Singapore (Bulluck et al., 2019) have been utilizing this method in their research to determine the mortality of patients with ACS. In Malaysia, there is only one study regarding the mortality prediction in ACS at 30 days using the ML method by Aziida et al. (2021) but there has no study yet reported on the ACS subtypes; STEMI and NSTEMI/UA specifically. The lack of a dataset applicable to the heterogeneous Malaysian population makes adaptation an inaccurate process and may impair patient care. Hence, a model based on data from Malaysians should be built to ensure that the mortality of the patient with STEMI and NSTEMI/UA can be predicted so that a well-prepared process can be executed by clinicians. The summary of the previous studies is summarised in Table 2.13 below.

Table 2.13: Summary of previous studies that incorporate ML in predicting mortality of ACS patients.

Country	Title	No. of instances	Input variables	Output variables	Performance metrics
United States of America	Determinants of In-Hospital Mortality After Percutaneous Coronary Intervention: A Machine Learning Approach	479 804 patients	49 variables	Ranking of variable importance	**AUC (95% CI)** **AdaBoost** 0.927(0.923–0.929) **XGBoost** 0.913 (0.906– 0.919)

Table 2.13, continued.

Country	Title	No. of instances	Input variables	Output variables	Performance metrics
	(Al'Aref et al., 2019)				**Logistic regression** 0.908 (0.907–0.910) **Random Forest** 0.892 (0.889–0.896)
United States of America	Building Computational Models to Predict One-Year Mortality in ICU Patients with Acute Myocardial Infarction and Post Myocardial Infarction Syndrome (Barrett et al., 2019)	5346 admissions	75 variables	No feature selection	**AdaBoost** AUC: 0.849 **Bayes Net** AUC: 0.744 **Decision Stump** AUC: 0.730 **Decision Table** AUC: 0.865 **J48** AUC: 0.843 **JRip** AUC: 0.737 **LMT** AUC: 0.901 **Logistic** AUC: 0.899

Table 2.13, continued.

Country	Title	No. of instances	Input variables	Output variables	Performance metrics
					Naïve Bayes AUC: 0.768 **OneR** AUC: 0.749 **PART** AUC: 0.869 **Random Forest** AUC: 0.893 **Random Tree** AUC: 0.776 **REP Tree** AUC: 0.845 **SGD** AUC: 0.765 **Simple Logistic** AUC: 0.901 **SMO** AUC:0.751

Table 2.13, continued.

Country	Title	No. of instances	Input variables	Output variables	Performance metrics
					Voted Perceptron AUC: 0.519 **Deep FNN** AUC: 0.751
United States of America	Machine Learning Improves Risk Stratification After Acute Coronary Syndrome (Myers et al., 2017)	4935 patients	8 variables	No feature selection	**AUC (95% CI)** **Logistic RegressionHX** 0.695 (0.581-0.809) **Logistic RegressionST** 0.701 (0.587-0.814) **Logistic RegressionHX+ST** 0.734 (0.623-0.845) **Logistic RegressionHX_MV** 0.727 (0.615-0.839) **Logistic RegressionHX+HRV** 0.720 (0.607-0.832) **Logistic RegressionHX+DC** 0.705 (0.591-0.818)

Table 2.13, continued.

Country	Title	No. of instances	Input variables	Output variables	Performance metrics
					TIMI 0.670 (0.555-0.786) **Recurrent Neural Network** 0.689 (0.575-0.803) **Artificial Neural Network** 0.743 (0.633-0.853
United States of America	Risk prediction model for in-hospital mortality in women with ST-elevation myocardial infarction: A machine learning approach (Mansoor et al., 2017)	12,047 patients	32 variables	11 variables for Logistic Regression 17 variables for Random Forest	**Logistic Regression (selected variables)** AUC: 0.85 **Random Forest (full model)** AUC: 0.82 **Random Forest (selected variables)** AUC: 0.81
United States of America	Prediction of 30-Day All-Cause Readmissions in Patients Hospitalized for Heart Failure Comparison of Machine Learning and Other Statistical Approaches (Frizzell et al., 2017)	56 477 patients	14 variables	No feature selection	**Tree-augmented naïve Bayesian network** AUC: 0.62 **Logistic regression** AUC: 0.62 **Least Absolute Shrinkage and**

Table 2.13, continued.

Country	Title	No. of instances	Input variables	Output variables	Performance metrics
					Selection Operator AUC: 0.62 **Gradient Boost model** AUC: 0.61 **Random Forest** AUC: 0.61 **EHR** AUC: 0.589
United States of America	Machine Learning for Risk Prediction of Acute Coronary Syndrome (VanHouten et al., 2014)	8408 records	88 variables	No feature selection	AUC (95% CI) **Random Forest** 0.848 (0.841-0.857) **Elastic Net** 0.818 (0.808-0.828) **Ridge Regression** 0.810 (0.801-0.820) **Modified TIMI** 0.745 (0.737-0.755) **Modified GRACE** 0.623 (0.615-0.634)

Table 2.13, continued.

Country	Title	No. of instances	Input variables	Output variables	Performance metrics
Sweden	Predicting two-year survival versus non-survival after first myocardial infarction using machine learning and Swedish national register data (Wallert et al., 2017)	51,943 patients	39 variables	Test with 39, 10 and 5 variables respectively	(39,10,5) Variables **LR** Sensitivity: (0.771, 0.754, 0.749) Specificity: (0.770, 0.758, 0.750) PPV: (0.293, 0.278, 0.270) NPV: (0.965, 0.961, 0.960) Detection Rate: (0.085, 0.0.83, 0.082) Detection incidence: (0.290, 0.298, 0.305) Accuracy: (0.770, 0.758, 0.750) **Boosted C5.0** Sensitivity: (0.798, 0.768, 0.758) Specificity: (0.739, 0.757, 0.736) PPV: (0.293, 0.278, 0.270) NPV: (0.965, 0.961, 0.960) Detection Rate: (0.088, 0.084, 0.083) Detection incidence: (0.320, 0.301, 0.319) Accuracy: (0.746, 0.758, 0.738) **RF** Sensitivity: (0.789, 0.771, 0.755) Specificity: (0.752, 0.746, 0.703) PPV: (0.282, 0.272, 0.239) NPV: (0.966, 0.963, 0.959) Detection Rate: (0.087, 0.085, 0.083)

Table 2.13, continued.

Country	Title	No. of instances	Input variables	Output variables	Performance metrics
					Detection incidence: (0.307, 0.311, 0.348) Accuracy: (0.756, 0.748, 0.708) **<u>SVM</u>** Sensitivity: (0.784, 0.751, 0.732) Specificity: (0.751, 0.756, 0.753) PPV: (0.280, 0.275, 0.268) NPV: (0.966, 0.961, 0.958) Detection Rate: (0.086, 0.083, 0.080) Detection incidence: (0.308, 0.300, 0.300) Accuracy: (0.755, 0.755, 0.751)
Poland	Predicting Long-Term Mortality after Acute Coronary Syndrome Using Machine Learning Techniques and Hematological Markers (Pieszko et al., 2019)	5053 patients	19 variables	No feature selection	**<u>In-hospital</u>** **Gradient-boosted tree** AUC: 0.89 **GRACE** AUC: 0.90 **<u>6 months</u>** **Gradient-boosted tree** AUC: 0.77 **GRACE** AUC: 0.73

Table 2.13, continued.

Country	Title	No. of instances	Input variables	Output variables	Performance metrics
					1-year **Gradient-boosted tree** AUC: 0.72 **GRACE** AUC: 0.71
Israel	Machine learning for prediction of 30-day mortality after ST-elevation myocardial infarction: An Acute Coronary Syndrome Israeli Survey data mining study (Shouval et al., 2017)	2782 patients	54 variables	Feature selection with all, 50, 40, 30, 20, 15, 10, 5 variables	**RF** AUC: 0.91 **Naïve Bayes** AUC: 0.87 **AdaBoost** AUC: 0.87 **LR** AUC:0.86 **Alternating Decision Tree (ADT)** AUC :0.84 **Pruning rules-based classification tree (PART)** AUC:0.64 **GRACE Score** AUC: 0.87 **TIMI Score** AUC: 0.82
China	Machine Learning to Predict the 1-Year Mortality Rate After Acute Anterior Myocardial Infarction in Chinese Patients	1244 patients	29 variables	Top 20 variables from the random forest and XGBoost	**Logistic Regression** AUC all: 0.931; AUC selected:0.864

Table 2.13, continued.

Country	Title	No. of instances	Input variables	Output variables	Performance metrics
	(Li et al., 2020)				**Gaussian Naïve Bayes** AUC all: 0.924; AUC selected: 0.909 **K Neighbors** AUC all: 0.709; AUC selected: 0.784 **Decision Tree** AUC all: 0.772; AUC selected: 0.852 **Random Forest** AUC all: 0.932; AUC selected: **XGBoost** AUC all: 0.942; AUC selected: 0.913
China	Machine Learning Models to Predict In-Hospital Mortality for St-Elevation Myocardial Infarction: From China Acute Myocardial Infarction (Cami) Registry (Li et al., 2017)	18744 patients	87 Variables	No feature selection	**Logistic Regression** AUC (CI): 0.860 (0.844-0.875) **Random Forest** AUC (CI): 0.868 (0.853-0.883) **Bayesian Network** AUC (CI): 0.861 (0.846-0.877)

Table 2.13, continued.

Country	Title	No. of instances	Input variables	Output variables	Performance metrics
					GRACE Score AUC (CI): (0.782, 0.763-0.800) **TIMI Score** AUC (CI): (0.807,0.788-0.826)
Korea	A soft voting ensemble classifier for early prediction and diagnosis of occurrences of major adverse cardiovascular events for STEMI and NSTEMI during 2-year follow-up in patients with acute coronary syndrome (Sherazi et al., 2021)	11,189 subjects (5389 STEMI, 5800 NSTEMI)	56 variables	Top 10 variable importance ranking using random forest, extra tree, and gradient boosting machine	**Overall dataset** **Random Forest** AUC: 0.990 **Extra tree** AUC: 0.995 **Gradient Boosting Machine** AUC: 0.989 **Soft Voting Ensemble** AUC: 0.996 **STEMI dataset** **Random Forest** AUC: 0.982 **Extra tree** AUC: 0.990

Table 2.13, continued.

Country	Title	No. of instances	Input variables	Output variables	Performance metrics
					Gradient Boosting Machine AUC: 0.993 **Soft Voting Ensemble** AUC: 0.995 **NSTEMI dataset** **Random Forest** AUC: 0.988 **Extra tree** AUC: 0.990 **Gradient Boosting Machine** AUC: 0.994 **Soft Voting Ensemble** AUC: 0.994
Korea	Machine learning enhances the performance of short and long-term mortality prediction models in non-ST-segment	- 14,183 subjects (5557 STEMI, 8626 NSTEMI)	46 variables	Feature ranking by variable importance	**STEMI dataset** **In-hospital** **Lasso** AUC: 0.890 **Ridge** AUC: 0.889

Table 2.13, continued.

Country	Title	No. of instances	Input variables	Output variables	Performance metrics
	elevation myocardial infarction (Lee et al., 2021)				**Elastic Net** AUC: 0.890 **Random Forest** AUC: 0.910 **Support Vector Machine** AUC: 0.819 **XGBoost** AUC: 0.912 **TIMI** AUC: 0.855 **GRACE** AUC: 0.896 **ACTION** AUC: 0.891 **3 months** **Lasso** AUC: 0.777 **Ridge** AUC: 0.779 **Elastic Net** AUC: 0.777 **Random Forest** AUC: 0.763 **Support Vector Machine** AUC: 0.667

Table 2.13, continued.

Country	Title	No. of instances	Input variables	Output variables	Performance metrics
					XGBoost AUC: 0.784 **TIMI** AUC: 0.743 **GRACE** AUC: 0.766 **ACTION** AUC: 0.709 **12 months** **Lasso** AUC: 0.835 **Ridge** AUC: 0.840 **Elastic Net** AUC: 0.835 **Random Forest** AUC: 0.825 **Support Vector Machine** AUC: 0.684 **XGBoost** AUC: 0.806 **TIMI** AUC: 0.793

Table 2.13, continued.

Country	Title	No. of instances	Input variables	Output variables	Performance metrics
					GRACE AUC: 0.826 **ACTION** AUC: 0.780 **NSTEMI dataset** **In-hospital** **Lasso** AUC: 0.886 **Ridge** AUC: 0.885 **Elastic Net** AUC: 0.886 **Random Forest** AUC: 0.889 **Support Vector Machine** AUC: 0.760 **XGBoost** AUC: 0.888 **TIMI** AUC: 0.669 **GRACE** AUC: 0.873

Table 2.13, continued.

Country	Title	No. of instances	Input variables	Output variables	Performance metrics
					ACTION AUC: 0.871 **3 months** **Lasso** AUC: 0.849 **Ridge** AUC: 0.826 **Elastic Net** AUC: 0.849 **Random Forest** AUC: 0.799 **Support Vector Machine** AUC: 0.715 **XGBoost** AUC: 0.824 **TIMI** AUC: 0.672 **GRACE** AUC: 0.777 **ACTION** AUC: 0.795 **12 months** **Lasso** AUC: 0.860

Table 2.13, continued.

Country	Title	No. of instances	Input variables	Output variables	Performance metrics
					Ridge AUC: 0.858 **Elastic Net** AUC: 0.859 **Random Forest** AUC: 0.836 **Support Vector Machine** AUC: 0.729 **XGBoost** AUC: 0.851 **TIMI** AUC: 0.675 **GRACE** AUC: 0.808 **ACTION** AUC: 0.790
Singapore	Independent Predictors of Cardiac Mortality and Hospitalization for Heart Failure in a Multi-Ethnic Asian ST-segment Elevation Myocardial Infarction Population Treated by Primary Percutaneous Coronary	11,546 patients	36 variables	7 variables for each of the time frame in-hospital, 30 days and 1-year	**Logistic regression** **AUC (95% CI)** **In-hospital** 0.921 (0.910-0.932) **30 days** 0.901 (0.887-0.915) **1-year** 0.881 (0.867-0.896)

Table 2.13, continued.

Country	Title	No. of instances	Input variables	Output variables	Performance metrics
	Intervention (Bulluck et al., 2019)				
Malaysia	Predicting 30-Day Mortality after an Acute Coronary Syndrome (ACS) using Machine Learning Methods for Feature Selection, Classification and Visualisation (Aziida et al., 2021)	302 patients	54 variables	The number of variables varies from different feature selection (Sequential backward elimination based on variable important for random forest, support vector machine with radial basis function, logistic regression and elastic net, cluster dendrogram, Boruta, recursive feature elimination, learning vector quantization, and genetic algorithm).	The best model for each of the ML method **RFimp-SBS-RF** AUC: 0.79 **RFE-SVM** AUC: 0.77 **GA-SBS-EN** AUC: 0.79 **LVQ-SBS-LR** AUC: 0.75 **TIMI Score** AUC: 0.60

All these studies reported that the ML method outperformed conventional risk scoring methods in predicting patient mortality. For an example, in the study by Pieszko et al.

(2019) on “Predicting Long-Term Mortality after Acute Coronary Syndrome Using Machine Learning Techniques and Hematological Markers”, for six-month mortality, the results of the best model and the GRACE score on the test set are 0.77 and 0.73, respectively while for 1-year mortality, the AUC for ML model is 0.72 and GRACE score of 0.71 indicating that ML performs better than the conventional GRACE score. The same situation was reported in the study by Shouval et al. (2017) in predicting mortality of patients with STEMI after 30 days. The AUC value of their best ML model, RF, with the value of 0.91 is higher compared to the conventional methods GRACE score and TIMI score with the value of 0.87 and 0.82 respectively. The ML models achieved significantly better performance in predicting in-hospital mortality than the traditional statistical analysis, which could improve the treatment decision for STEMI patients. This provides the insight that ML might be the better way to build a prediction model.

However, most of the past studies in Table 2.13 are mostly for ACS or STEMI patients and it limits the use of the models and the target groups. Only recently, Sherazi et al. (2021) and Lee et al. (2021) researched NSTEMI/UA patients using ML approaches to build a prediction model. Both studies still display a good performance by their best ML methods where the AUC for NSTEMI mortality prediction was still high in both studies with the value of 0.99 and 0.89 respectively.

To determine the best algorithm for model building, a comparison must be made among the ML methods and the conventional statistical method. A study on the 30-days readmission of hospitalized patients for heart failure (Frizzell et al., 2017) used several ML algorithms to test their prediction. NB, RF, Gradient boosted (GB) and LR were adopted for the model building. The results showed that all the ML methods' AUC values are almost similar to each other; 0.62, 0.61, 0.61, and 0.62, and the conventional statistical method

with only a 0.59 AUC value. According to the study, this set of ML methods are deployable for the new study although the result is almost the same which requires a comparison among them so that the best model could be selected and later be deployed to a working system.

Based on the list of ML methods above, another study had been conducted to compare the performance of each of the ML methods. A study by Wallert et al. (2017) proved that SVM outperformed Boosted C5.0 but not significantly higher than LR or RF. This statement strengthens the reason for RF and SVM to be used as their performance is among the highest in comparison with the others. Apart from that, the study also mentioned that the model is preferable due to its vast coverage of sample and the attributes are significant to the population of a specific place, resulting in the model being more accurate.

Bulluck et al. (2019) used ML algorithms to predict characteristics linked to cardiac mortality in a multi-ethnic South-East Asian population, resulting in a strong model performance for short and long-term prediction. However, just one ML method (LR) was used in this work, which is identical to the traditional method of constructing the existing risk score for ACS mortality prediction. Overall, this study demonstrated that a multi-ethnic South-East Asian population study on ACS mortality prediction is feasible to implement and further test with many additional ML methods to reach the best prognostic model performance.

Six of the studies reported in Table 2.13 did not use feature selection to determine the variables that are essential in predicting the mortality prediction model for ACS patients (Barrett et al., 2019; Frizzell et al., 2017; Li et al., 2017; Myers et al., 2017; Pieszko et al., 2019; VanHouten et al., 2014) because, during the development of the model, these studies stated that they did not wish to lose any information by removing any variables. Bulluck et

al. (2019), Al'Aref et al. (2019), Sherazi et al. (2021), and Lee et al. (2021), on the other hand, solely ranked the variables based on variable importance to discuss the factors that are critical to their full models' performance. Whereas, the rest of the studies, used their selected variables from the feature selection method to develop models and compared their results to the full models and the traditional risk score. According to Mansoor et al. (2017) and Li et al. (2020), the performance of models with all and selected variables is nearly identical, however Shouval et al. (2017) and Wallert et al. (2017) found that models with all variables performed better than models with selected variables. Only Aziida et al. (2021) found that employing the feature selection technique improved the performance of mortality prediction using ML with higher AUC values. These studies suggest that the feature selection approach is significant to include in this study, even though the outcomes varied depending on the dataset used.

Amid the highly demanding application of ML in mortality prediction, a new method is emerging. DL, a sub-domain of ML, has recently established an interesting new trend in ACS prediction. The theoretical foundations are well-established in the literature on neural networks. DL provides for the utilisation of deep architectural advantage (multi-layered) combined with novel training paradigms, which are the differences (Benjamins et al., 2019; Ravì et al., 2016). DL, in a nutshell, is defined as a multi-layered neural network design. Table 2.14 summarises the research that has been conducted utilising DL to predict the mortality of ACS patients.

Table 2.14: Summary of previous studies that incorporates DL in predicting mortality of ACS patients.

Country	Title	No. of instances	Input variables	Output variables	Performance's metrics
Korea	A machine learning–based 1-year mortality prediction model after hospital discharge for clinical patients with acute coronary syndrome (Sherazi et al., 2020)	8227	69 variables (refer to the paper)	9 variables (top nine primary prognostic factors according to each machine learning models) (Refer to the paper)	**DNN** AUC: 0.898 Precision: 0.977 Recall: 0.927 Accuracy: 0.911 F-score: 0.951 **GBM Gradient Boosting Machine** AUC: 0.898 Precision: 0.967 Recall: 0.977 Accuracy: 0.947 F-score: 0.972 **GLM Generalized Linear Model** AUC: 0.873 Precision: 0.972 Recall: 0.949 Accuracy: 0.926 F-score: 0.960 **RF** AUC: 0.883 Precision: 0.976 Recall: 0.954 Accuracy: 0.935 F-score: 0.965 **GRACE** AUC: 0.810 Precision: 0.970 Recall: 0.922 Accuracy: 0.900 F-score: 0.946
Korea	Deep learning–based prediction model of occurrences of major adverse cardiac events during 1-year follow-up after hospital discharge in patients with AMI using knowledge mining	10813 subjects	49 variables	8 variables (Top eight primary prognostic factors according to each machine learning models) (Refer to the paper)	**(1M, 6M, 12M) DNN** Accuracy: (95.98, 95.28, 95.43) Sensitivity: (81.25. 71.43, 88.89) Specificity: (96.10, 95.44, 95.46) AUC: (0.97, 0.94, 0.96) **Gradient Boosting Machine GBM** Accuracy: (95.80, 95.15, 81.82)

Table 2.14, continued.

Country	Title	No. of instances	Input variables	Output variables	Performance's metrics
	(Kim et al., 2019)				Sensitivity: (68.75, 57.14, 96.44) Specificity: (96.00, 95.40, 96.36) AUC: (0.96, 0.95, 0.96) **Generalized linear model GLM** Accuracy: (93.69, 87.11, 91.96) Sensitivity: (22.50, 17.52, 16.56) Specificity: (96.97, 95.24, 97.32) AUC: (0.76, 0.67, 0.72) **GRACE** Accuracy: (81.51,89.34, 94.93) Sensitivity: (81.36, 25.71, 76.25) Specificity: (82.88, 92.51, 97.16) AUC: (0.75, 0.72, 0.76)
Korea	Deep-learning-based risk stratification for mortality of patients with acute myocardial infarction (Kwon, Jeon, et al., 2019a)	22857	36 variables	13 variables - Age - Sex - BMI - LDL - Heart Rate - CPR - Killip class - OHCA - Creatinine - Glucose - SBP - ST-Elevation - CK-MB	**STEMI** **Deep Learning (MLP)** AUC: 0.905 (95%CI): (0.902-0.909) **RF** AUC:0.890 (95%CI): (0.886-0.895) **LR** AUC:0.873 (95%CI): (0.869-0.878) **GRACE** AUC: 0.851 (95%CI): (0.846-0.856) **ACTION** AUC:0.852 (95%CI): (0.847-0.857)

Table 2.14, continued.

Country	Title	No. of instances	Input variables	Output variables	Performance's metrics
					TIMI AUC:0.781 (95%CI): (0.775-0.787) **NSTEMI** **Deep Learning (MLP)** AUC: 0.870 (95%CI): (0.865-0.876) **RF** AUC:0.851 (95%CI): (0.845-0.858) **LR** AUC:0.845 (95%CI): (0.839-0.851) **GRACE** AUC: 0.810 (95%CI): (0.803-0.819) **ACTION** AUC:0.806 (95%CI): (0.799-0.814) **TIMI** AUC:0.593 (95%CI): (0.5850.603)
Korea	Artificial intelligence algorithm for predicting mortality of patients with acute heart failure (Kwon et al., 2019b)	12,654	22 variables (Refer to the paper)	-	**Deep Learning (MLP)** AUC: 0.880 (95%CI): (0.876-0.884) **RF** AUC:0.756 (95%CI): (0.749-0.766) **LR** AUC:0.720 (95%CI): (0.712-0.730) **SVM** AUC: 0.723 (95%CI): (0.714-0.732) **Bayesian network** AUC:0.730

Table 2.14, continued.

Country	Title	No. of instances	Input variables	Output variables	Performance's metrics
					(95%CI): (0.721-0.739) **GWTG-HF score** AUC:0.728 (95%CI): (0.720-0.737)

In the studies carried out by Sherazi et al. (2020), Kim et al. (2019), and Kwon et al. (2019a), DL models using MLP/DNN outperformed conventional methods by achieving higher AUC.

For example, the research was done by Kwon et al. (2019a) showed that DL achieved the highest AUC of 0.905 as compared to conventional risk score GRACE (AUC=0.85), ACTION (AUC=0.85), and TIMI (AUC= 0.78) for STEMI patients risk prediction. Similarly, DL is the best predictive model for NSTEMI patients with an AUC of 0.870 as compared to GRACE (AUC=0.81), ACTION (AUC=0.81), and LR (AUC= 0.59).

Another example can be seen in the research done by Sherazi et al. (2020), where their deep neural network (AUC=0.90) outperformed the conventional risk score GRACE (AUC=0.81).

These performances of DL indicate that a good mortality prediction model for patients with ACS can be employed in the future. Hence, in this study, the DL method was used as an additional method to the ML method in predicting mortality and can be tested against the conventional score as the DL method is lacking in the feature selection component (Kwon et al., 2019a).

Together with statistical modeling and computer science, models of prediction of patient mortality after ACS can be built using data from heart disease patients. Further studies and

approaches can be conducted regarding the predictors of ACS mortality so that the disease could be prevented or stopped in the future.

CHAPTER 3: METHODOLOGY

3.1 Introduction

All the methods and materials used in this study are detailed in this chapter. As indicated in the Figure 3.2 below, this study involved different stages, including:

- data collecting
- data splitting
- data preprocessing
- model development
- parameter tuning
- feature selection,
- performance evaluation

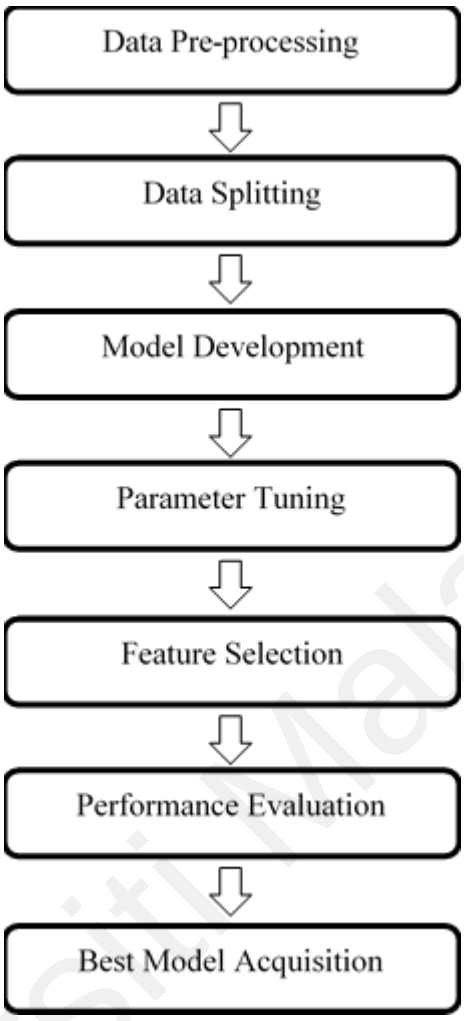


Figure 3.1: The general flowchart of the study.

3.2 Data collection

This study used data gathered between 2006 and 2016 from the Malaysian National Cardiovascular Database (NCVD-ACS) registry. The Medical Review & Ethics Committee (MREC) of Malaysia's Ministry of Health (MOH) approved the NCVD registry in 2007. (Approval Code: NMRR-07-20-250). MREC waived informed patient consent for NCVD. For consecutive patients treated at participating institutions, the registry collects data on a

standardised set of clinical, demographic, and procedural variables, as well as outcomes (Ahmad et al., 2011). The UiTM ethics committee (Reference number: 600-TNCPI (5/1/6) and the National Heart Association of Malaysia (NHAM) both gave their approval for data collection. NCDV-ACS is a Malaysian multicenter registry that includes up to 23 hospitals (Appendix A). Deaths were confirmed on a yearly basis through record connections with the Malaysian National Registration Department of deaths.

Data were collected using a standardised case report from the time ACS patients were admitted to the hospital until they discharged from the hospital and the follow-up afterward, along with the patients' outcome, which is alive or dead. The data included patient's demographic, clinical presentation, baseline investigation, electrocardiography, treatment, and pharmacological therapy. A unique identification number was assigned to each patient to prevent any duplication.

All patients from the ACS registry without exclusion were used including patients who received reperfusion such as primary PCI, angiography demonstrating spontaneous reperfusion, additional fibrinolysis, or urgent CABG. In this study, STEMI is defined as persistent ST-segment elevation ≥ 1 mm in two contiguous electrocardiographic leads, or the presence of a new left bundle branch block in the setting of positive cardiac markers with ischaemic type chest pain more than 30 minutes while NSTEMI is a prolonged ischaemic type chest pain with a non-interpretable resting ECG (such as paced rhythm or new bundle branch block) without ST elevation and UA is defined as ACS with myocardial ischemia but no observable myocardial necrosis (i.e., cardiac biomarkers of myocardial necrosis, such as creatine kinase MB isozyme, troponin, and myoglobin, are not discharged into the circulation) (MOH, 2011, 2019). NSTEMI and UA are closely related disorders

with comparable pathophysiologic origins and clinical manifestations, which is why patients with both conditions are grouped as NSTEMI/UA (Kumar & Cannon, 2009).

Based on clinical recommendations, 54 variables from a complete set of data for the STEMI study and 50 variables for NSTEMI/UA were used in this study. Categories of variables used were sociodemographic characteristics, CVD diagnosis and severity, CVD risk factors, CVD comorbidities, non-CVD comorbidities, biomarkers, and medication used. The mortality time points for in-hospital, 30 days, and 1-year were determined from the first hospital admission. Confirmation of deaths was done yearly via record linkages with the Malaysian National Registration Department. The follow-up data points are meant to collect these variables but unfortunately are excessive in terms of missing values and hence were omitted from the study. This study focused on the algorithm to policy changing endpoints for example hard endpoints such as death to increase the impact of the study. This is similarly done in other publications (Kwon et al., 2019a; Peng et al., 2017; Shouval et al., 2017).

3.3 Data Pre-processing

Data cleaning was done to remove the patients' records with missing data. A total of 33378 cases from the registry were collected and 12368 cases were identified as complete cases (without missing values on variables) for the STEMI study. Out of the 12368 cases, 6299 cases are in-hospital, 3130 cases are for 30 days and 2939 cases are for a 1-year mortality prediction study. As for the NSTEMI/UA study, a total of 42683 cases were identified and only 9,477 cases are complete cases. From the total number of complete cases, the number of cases for in-hospital, 30-days, and 1-year studies are 4771, 2402, and 2304 respectively. There are no missing values on the output data (alive/dead). Any

variables that have more than 50% of missing values were deleted beforehand and not considered in the study. The summary of the data components is tabulated in Table 3.1 below.

Table 3.1:The number of cases for STEMI and NSTEMI/UA study for each of the time points.

	In-hospital	30-days	1-year
STEMI			
Raw data	17,296	8,261	7,821
Records with missing outcome	10,997	5,131	4,882
Data with complete cases	6,299	3,130	2,939
NSTEMI/UA			
Raw data	23,809	9,774	9,100
Records with missing outcome	19,038	7,372	6,796
Data with complete cases	4,771	2,402	2,304

The initial total of variables selected by the expert clinicians from the registry was 54 variables for the STEMI dataset and 50 variables for NSTEMI/UA dataset. Each of the variables was then tested and ones with near-zero-variance or zero-variance values were

excluded from the dataset before being trained. This is to avoid any errors and unexpected results when sampling (Brownlee, 2020; Kuhn & Johnson, 2019). The variables that were being excluded from the whole STEMI dataset are:

- Peripheral vascular disease,
- ECG abnormalities type not-specific,
- ECG abnormalities type none,
- ECG abnormalities location none.

The variables that were excluded for the NSTEMI/UA dataset due to its near-to-zero or zero-variance values are:

- Peripheral vascular disease,
- ST-segment elevation≥1 mm (0.1 mV) in ≥2 contagious limb leads,
- ST-segment elevation≥2 mm (0.2mV) in ≥2 contagious limb leads or chest leads,
- ECG abnormalities location True posterior: V1, V2,
- ECG abnormalities location Right ventricle: ST elevation in lead V4R,
- Coronary artery bypass grafting (CABG) treatment,
- Aspirin intake,
- GP receptor inhibitor intake,
- Heparin intake,
- Other lipid-lowering agent intake,
- Anti-arrhythmic agent intake.

This rendered a full variable set of 50 variables for each time frame (9 continuous, 41 categorical) for the STEMI dataset and 39 variables (8 continuous, 31 categorical) for the

NSTEMI/UA dataset. Since a medical dataset was employed in this study, it was necessary to use the complete set of data in order to maintain data integrity and trustworthiness. However, data imputation was also carried out but further discussed later in this chapter in section 3.8. The percentage of the missing values in the remaining variables was tabulated in Table 3.2 (for STEMI) and Table 3.3 (for NSTEMI/UA).

Table 3.2: The percentage of missing values based on each variable for the STEMI dataset.

Variables	Attributes	In-hospital		30-days		1-year	
		Total missing value	Percentage of missing value (%)	Total missing value	Percentage of missing value (%)	Total missing value	Percentage of missing value (%)
Age		0	0.00	0	0.00	0	0.00
Race		0	0.00	0	0.00	0	0.00
Gender		0	0.00	0	0.00	0	0.00
Smoking status		590	3.41	346	4.19	332	4.24
History of hypertension		2039	11.79	1176	14.24	1145	14.64
History of diabetes		2223	12.85	1247	15.10	1214	15.52
Family history of premature cardiovascular disease		3135	18.13	1714	20.75	1641	20.98
History of myocardial infarction		2095	12.11	1041	12.60	999	12.77
Documented CAD		2468	14.27	1180	14.28	1131	14.46
History of heart failure		1851	10.70	940	11.38	898	11.48
Chronic lung disease		1880	10.87	953	11.54	912	11.66

Table 3.2, continued.

Variables	Attributes	In-hospital		30-days		1-year	
		Total missing value	Percentage of missing value (%)	Total missing value	Percentage of missing value (%)	Total missing value	Percentage of missing value (%)
Chronic renal disease		1854	10.72	930	11.26	894	11.43
Cerebrovascular disease		1825	10.55	926	11.21	882	11.28
Heart rate		554	3.20	214	2.59	200	2.55
Systolic blood pressure		455	2.63	140	1.69	129	1.65
Diastolic blood pressure		562	3.25	137	1.66	123	1.57
Killip class		679	3.93	235	2.84	207	2.65
Total cholesterol		2667	15.42	1462	17.70	1416	18.11
HDL		2950	17.06	1559	18.87	1510	19.31
LDL		2947	17.04	1554	18.81	1471	16.87
Triglycerides		2710	15.67	1531	18.53	1438	18.39
Fasting blood glucose		2890	16.71	1610	19.49	1521	19.45

Table 3.2, continued.

Variables	Attributes	In-hospital		30-days		1-year	
		Total missing value	Percentage of missing value (%)	Total missing value	Percentage of missing value (%)	Total missing value	Percentage of missing value (%)
ECG abnormalities type	ST segment elevation ≥1mm in ≥2 contiguous limb leads	0	0.00	0	0	0	0.00
	ST segment elevation ≥2mm in ≥2 contiguous frontal leads or chest leads	0	0.00	0	0	0	0.00
	ST segment depression ≥0.5mm in ≥2 contiguous leads	0	0.00	0	0	0	0.00
	T-wave inversion ≥1mm	0	0.00	0	0	0	0.00
	Bundle branch block	0	0.00	0	0	0	0.00
ECG abnormalities location	Inferior leads: II, III, aVF		0.00		0	0	0.00
	Anterior leads: V1 to V4	0	0.00	0	0	0	0.00

Table 3.2, continued.

Variables	Attributes	In-hospital		30-days		1-year	
		Total missing value	Percentage of missing value (%)	Total missing value	Percentage of missing value (%)	Total missing value	Percentage of missing value (%)
	Lateral leads: I, aVL, V5 to V6	0	0.00	0	0	0	0.00
	True posterior: V1, V2	0	0.00	0	0	0	0.00
	Right ventricle: ST elevation in lead V4R	0	0.00	0	0	0	0.00
FB status		253	1.46	56	0.68	44	2.57
Cardiac catheterization		605	3.50	206	2.49	201	7.25
PCI		1262	7.30	588	7.12	567	0.56
CABG		1498	8.66	606	7.34	591	7.56
ASA		337	1.95	144	1.74	133	1.70
GP receptor inhibitor		2321	13.42	1195	14.47	1127	14.41
Heparin		2189	12.66	1129	13.67	1063	13.59
LMWH		2066	11.94	1098	13.29	1045	13.36
Beta blockers		1604	9.27	882	10.68	828	10.59

Table 3.2, continued.

Variables	Attributes	In-hospital		30-days		1-year	
		Total missing value	Percentage of missing value (%)	Total missing value	Percentage of missing value (%)	Total missing value	Percentage of missing value (%)
ACE inhibitors		1688	9.76	891	10.79	840	10.74
Angiotensin II receptor blocker		2336	13.51	1200	14.53	1128	14.42
Statin		490	2.83	205	2.48	185	2.37
Other lipid lowering agent		2284	13.21	1197	14.49	1125	14.38
Diuretics		2012	11.63	1043	12.63	988	12.63
Calcium antagonist		2318	13.40	1200	14.53	1128	14.42
Oral hypoglycaemic agent		2161	12.49	1145	13.86	1078	13.78
Insulin		1954	11.30	1016	12.30	968	12.38
Anti-arrhythmic agent		2343	13.55	1233	14.93	1172	14.99

Table 3.3: The percentage of missing values based on each variable for NSTEMI/UA dataset.

Variables	Attributes	In-hospital		30-days		1-year	
		Total missing value	Percentage of missing value (%)	Total missing value	Percentage of missing value (%)	Total missing value	Percentage of missing value (%)
Age		0	0.00	0	0.00	0	0.00
Race		0	0.00	0	0.00	0	0.00
Gender		0	0.00	0	0.00	0	0.00
Smoking status		1573	6.61	631	6.46	571	6.27
History of dyslipidaemia		3731	15.67	1108	11.34	1021	11.22
History of hypertension		2252	9.46	826	8.45	779	8.56
History of diabetes		1613	6.77	588	6.02	559	6.14
Family history of premature cardiovascular disease		5909	24.82	1911	19.55	1706	18.75
History of MI/CAD		2078	8.73	692	7.08	637	7.00
New onset angina (<2 weeks)		2274	9.55	818	8.37	757	8.32
History of heart failure		3021	12.69	1209	12.37	1127	12.38

Table 3.3, continued.

Variables	Attributes	In-hospital		30-days		1-year	
		Total missing value	Percentage of missing value (%)	Total missing value	Percentage of missing value (%)	Total missing value	Percentage of missing value (%)
Chronic lung disease		3057	12.84	1199	12.27	1115	12.25
Chronic renal disease		3176	13.34	1144	11.70	1065	11.70
Cerebrovascular disease		3304	13.88	1191	12.19	1110	12.20
Heart rate		751	3.15	271	2.77	229	2.52
Systolic blood pressure		707	2.97	209	2.14	172	1.89
Diastolic blood pressure		857	3.60	224	2.29	182	2.00
Killip class		5474	21.82	3614	36.98	3380	37.14
Total cholesterol		7456	31.32	2642	27.03	2360	25.93
HDL		7418	31.16	2714	27.77	2443	26.85
LDL		7567	31.78	2710	27.73	2431	26.71
Fasting blood glucose		7664	32.19	2849	29.15	2577	28.32

Table 3.3, continued.

Variables	Attributes	In-hospital		30-days		1-year	
		Total missing value	Percentage of missing value (%)	Total missing value	Percentage of missing value (%)	Total missing value	Percentage of missing value (%)
ECG abnormalities type	ST-segment depression ≥0.5mm in ≥2 contiguous leads	0	0.00	0	0.00	0	0.00
	T-wave inversion ≥1mm	0	0.00	0	0.00	0	0.00
	Bundle branch block	0	0.00	0	0.00	0	0.00
ECG abnormalities location	Inferior leads: II, III, aVF	0	0.00	0	0.00	0	0.00
	Anterior leads: V1 to V4	0	0.00	0	0.00	0	0.00
	Lateral leads: I, aVL, V5 to V6	0	0.00	0	0.00	0	0.00
Cardiac catheterization		1121	4.71	298	3.05	272	2.99
PCI		1643	6.90	524	5.36	489	5.37
LMWH		2096	8.80	1097	11.22	1018	11.19
Beta-blockers		1359	5.71	532	5.44	470	5.16

Table 3.3, continued.

Variables	Attributes	In-hospital		30-days		1-year	
		Total missing value	Percentage of missing value (%)	Total missing value	Percentage of missing value (%)	Total missing value	Percentage of missing value (%)
ACE inhibitors		1580	6.64	668	6.83	589	6.47
Angiotensin II receptor blocker		2468	10.37	1161	11.88	1061	11.66
Statin		828	3.48	250	2.56	207	2.27
Diuretics		2102	8.83	1028	10.52	944	10.37
Calcium antagonist		2343	9.84	1111	11.37	1006	11.05
Oral hypoglycaemic agent		2114	8.88	996	10.19	905	9.95
Insulin		2226	9.35	1074	10.99	988	10.86

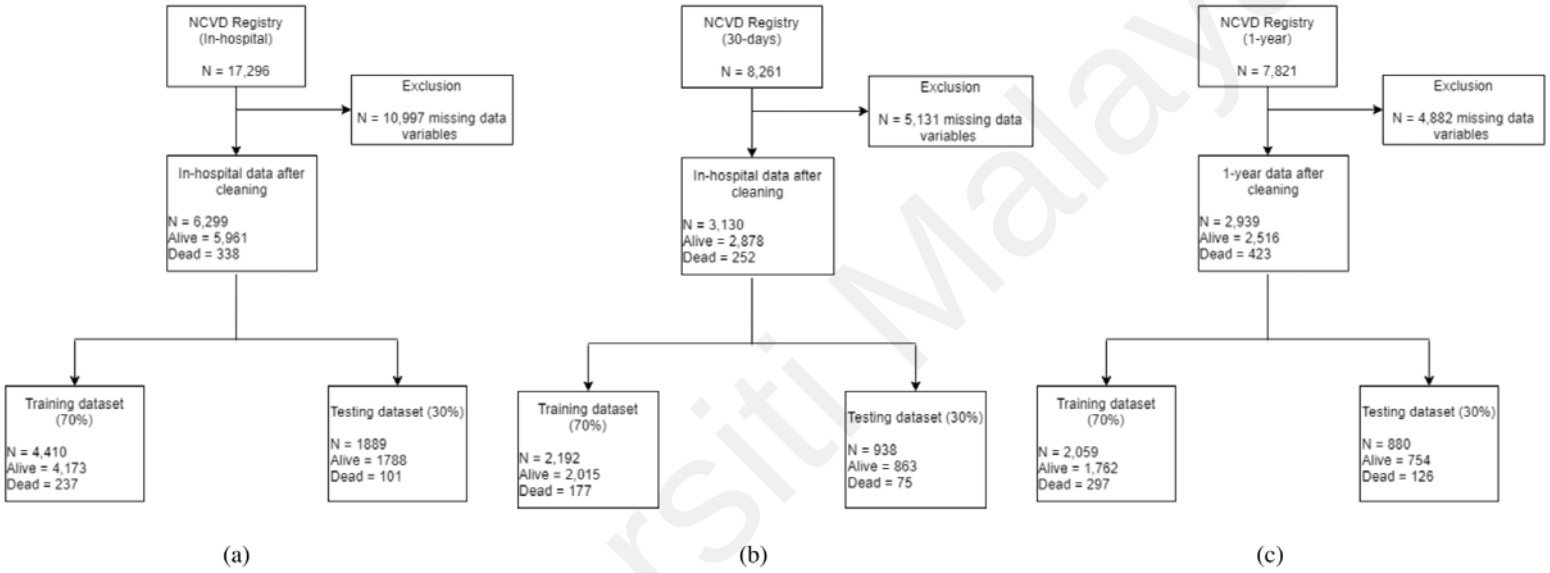


Figure 3.2: The flow chart indicating the numbers of instances before and after data cleaning in STEMI dataset for (a) in-hospital, (b) 30-days, and (c) 1-year.

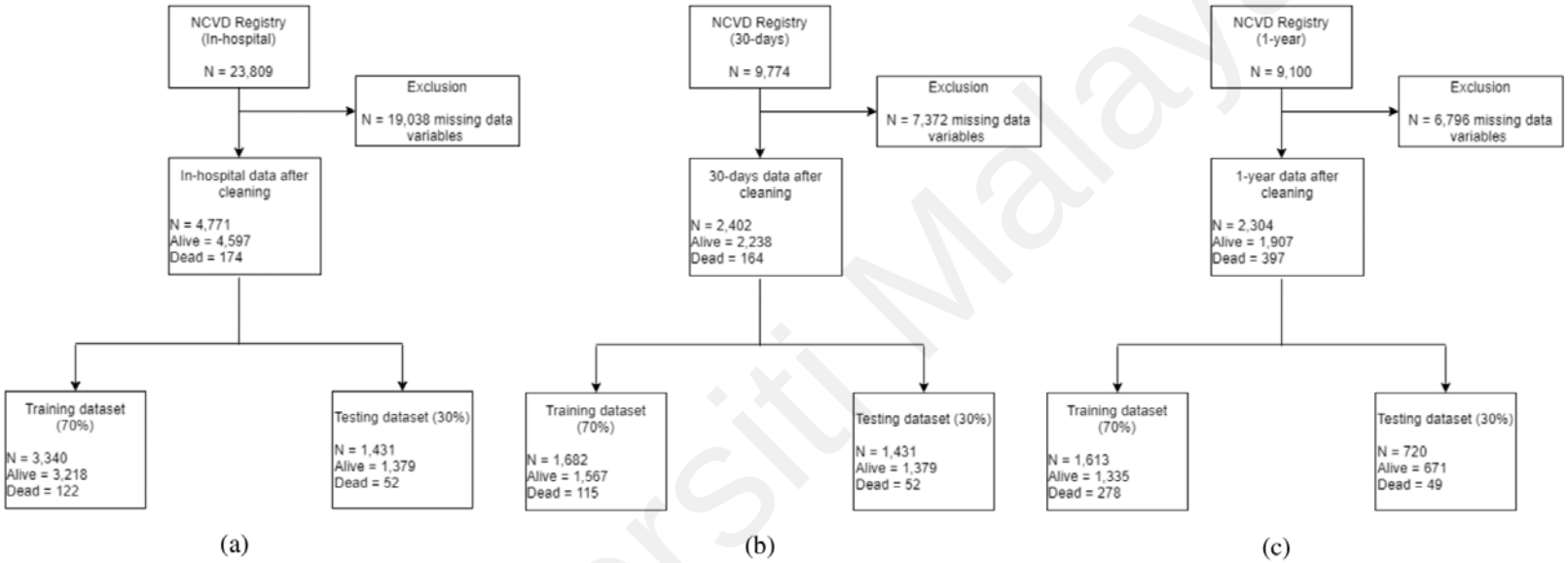


Figure 3.3:The flow chart indicating the numbers of instances before and after data cleaning in NSTEMI/UA dataset for (a) in-hospital, (b) 30-days, and (c) 1-year.

From the raw data to the finalised training and testing data used in each of the model building, Figure 3.2 and Figure 3.3 summarise the data cleaning process for both STEMI and NSTEMI/UA datasets. The information and data types of the variables of ACS data acquired from the registry are shown in Table 3.4.

Table 3.4: The variables used in the study, their data domain, and data types for both STEMI and NSTEMI/UA datasets.

Variables		Data domain	Data type
Demographic			
Age		≥20	Continuous
Race		1: Malay 2: Chinese 3: Indian 4: Others	Categorical
Gender		1: Male 2: Female	Categorical
Status Before Event			
Smoking status		1: Never 2: Quit (>30 days) 3: Current	Categorical
History of dyslipidaemia		1: Yes 2: No	Categorical
History of hypertension		1: Yes 2: No	Categorical
History of diabetes		1: Yes 2: No	Categorical

Table 3.4, continued.

Variables		Data domain	Data type
Family history of premature cardiovascular disease		1: Yes 2: No	Categorical
History of myocardial infarction		1: Yes 2: No	Categorical
Documented CAD		1: Yes 2: No	Categorical
New onset angina (<2 weeks)		1: Yes 2: No	Categorical
History of heart failure		1: Yes 2: No	Categorical
Chronic lung disease		1: Yes 2: No	Categorical
Chronic renal disease		1: Yes 2: No	Categorical
Cerebrovascular disease		1: Yes 2: No	Categorical
Clinical Presentation and Examination			
Heart rate		20 - 200 beats/min	Continuous
Systolic blood pressure		50 - 270 mmHg	Continuous
Diastolic blood pressure		10 - 170 mmHg	Continuous
Killip class		1: Killip class I 2: Killip class II 3: Killip class III	Categorical

Table 3.4, continued.

Variables		Data domain	Data type
		4: Killip class IV	
Baseline investigation			
Total cholesterol		2.0 - 25.0 mmol/L	Continuous
HDL		0.5 – 5.0 mmol/L	Continuous
LDL		0.5 – 20.0 mmol/L	Continuous
Triglycerides		0.5 – 15.0 mmol/L	Continuous
Fasting blood glucose		3.0 -50.0 mmol/L	Continuous
Electrocardiography			
ECG abnormalities type	ST-segment elevation ≥1mm in ≥2 contiguous limb leads	1: Yes 2: No	Categorical
	ST-segment elevation ≥2mm in ≥2 contiguous frontal leads or chest leads	1: Yes 2: No	Categorical
	ST-segment depression ≥0.5mm in ≥2 contiguous leads	1: Yes 2: No	Categorical
	T-wave inversion ≥1mm	1: Yes 2: No	Categorical
	Bundle branch block	1: Yes 2: No	Categorical
ECG abnormalities location	Inferior leads: II, III, aVF	1: Yes 2: No	Categorical
	Anterior leads: V1 to V4	1: Yes 2: No	Categorical
	Lateral leads: I, aVL, V5 to V6	1: Yes 2: No	Categorical

Table 3.4, continued.

Variables		Data domain	Data type
	True posterior: V1, V2	1: Yes 2: No	Categorical
	Right ventricle: ST elevation in lead V4R	1: Yes 2: No	Categorical
Invasive Therapeutic Procedures			
FB status		1: Yes 2: No	Categorical
Cardiac catheterization		1: Yes 2: No	Categorical
PCI		1: Yes 2: No	Categorical
CABG		1: Yes 2: No	Categorical
Pharmacological Therapy (Medication)			
ASA		1: Yes 2: No	Categorical
GP receptor inhibitor		1: Yes 2: No	Categorical
Heparin		1: Yes 2: No	Categorical
LMWH		1: Yes 2: No	Categorical
Beta-blockers		1: Yes 2: No	Categorical
ACE inhibitors		1: Yes 2: No	Categorical

Table 3.4, continued.

Variables		Data domain	Data type
Angiotensin II receptor blocker		1: Yes 2: No	Categorical
Statin		1: Yes 2: No	Categorical
Other lipid-lowering agent		1: Yes 2: No	Categorical
Diuretics		1: Yes 2: No	Categorical
Calcium antagonist		1: Yes 2: No	Categorical
Oral hypoglycaemic agent		1: Yes 2: No	Categorical
Insulin		1: Yes 2: No	Categorical
Anti-arrhythmic agent		1: Yes 2: No	Categorical

3.4 R Packages and Library

R (R Development Core Team, Vienna, Austria) was the main language that was used throughout the study which included the process of data pre-processing and the models' development. The STEMI ML models were developed with R package version 3.5.2, while the NSTEMI/UA ML models, STEMI DL models, and NSTEMI/UA DL models were developed with R package version 4.0.3. Table 3.5 below summarises all the R libraries/packages that we used for the study (Bunn & Korpela, 2020).

Table 3.5: R library and packages used in this study.

Libraries/Packages	Functions
readr	Provide a quick and easy way to read various rectangular text data types like csv and tsv.
dplyr	It's a data manipulation grammar for dealing with a data frame-like object.
ROSE	The Random Over-Sampling Examples package contains functions for dealing with the occurrence of imbalanced classes in binary classification tasks.
keras	Keras, which represents the high-level neural network API, has an R interface. Keras provides rapid neural network experimentation.
Tenserflow	It's an R interface to TensorFlow, an open-source software framework for numerical computations based on data flow graphs. As a backend for the creation of MLP models.
Caret	Classification and Regression Training is what this acronym stands for. It is a software program that has many capabilities that will aid in the creation of prediction models. Data splitting, data pre-processing, feature selection, model tuning, and variable importance estimation are some of the tools contained in the package.
ROCR	Use to visualise the scoring classifiers' performance. Use to generate ROC graphs.
pROC	Use for visualizing, smoothing, and comparing the ROC curves.

3.5 Data Splitting

For the ML models, data was split into 70% training data and 30% testing data (as can be seen in Figures 3.2 and 3.3 above) to precisely determine the performance of prediction models. Rather than splitting the data into training, validation, and test subsets and performing holdout cross-validation on the validation set, the data was split into training and testing sets, and k-fold cross-validation was performed, which is a better method than

the traditional train-validation-test split cross-validation (Hsieh et al., 2019). Both groups were subjected to 10-fold cross-validation at three different time points during the prediction model development process. The data was split into 70:30 parts for several reasons:

a) Proven from the previous study

Gholamy et al. (2018) found that allocating 70% - 80% of original data for training and the remaining 20% - 30% for testing yields the best outcomes for several empirical results.

b) Size of the dataset

Since the dataset for both STEMI and NSTEMI/UA dataset for all three time points are approximately between ~2000 to ~6000 cases (near to ~10000), splitting the data into 70:30 is suggested as the optimum division by Gholamy et al. (2018). The suggested splitting division is shown in Figure 3.4 below.

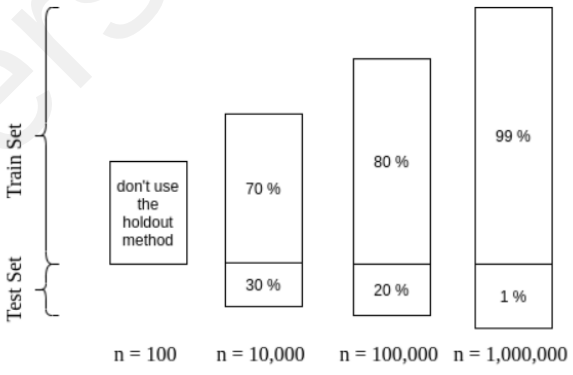


Figure 3.4: The suggested optimal ratio of the train-test split according to the size of the dataset (Gholamy et al., 2018).

c) The variance affecting the performance of the model

The parameter estimations have a significant variance when there is less training data. Less testing data, on the other hand, leads to a lot of variation in performance measurements. It may lead to inaccuracy of the result (Mehta et al., 2019).

3.6 Data Balancing

One of the requirements of ML is that the data set must be balanced, or at least near to being balanced. The fundamental reason for this is to ensure that each class receives equal attention (Lunardon et al., 2014). As can be seen in Figures 3.2 and 3.3, the training data for each of the datasets for the three time points have an imbalanced collection of samples in two classes: dead (0) and alive (1). Unsatisfactory classifications will result from the huge data imbalance. Furthermore, data balancing enabled us to give equal weight to both the dead and the alive classes. The purpose of data balancing was to eliminate bias in favour of the majority class, which would result in poor classification of the minority class.

The data balancing method was carried out using the ROSE (Random Over Sampling Examples) tool in R. ROSE is a binary imbalanced learning software. It aids in the generation of synthetic data using sampling methods and a smoothed bootstrap approach. This package has well-defined accuracy functions for quickly completing tasks. It can also deal with both continuous and categorical data. There are three types of resampling techniques that were considered and tested for this study using the ROSE package; the "undersampling" method, "oversampling" method, and "both under- and over-sampling methods combined". All three techniques were tested on each dataset, with the "both under- and over-sampling methods combined" method proving to be the most effective, it was applied to all the STEMI and NSTEMI/UA datasets for the data balancing stage in the data preparation of this study.

3.7 Data normalisation

The goal of data normalisation is to convert the values of continuous variables in a dataset to a common scale while preserving disparities in value ranges. Another scaling strategy is standardisation, or z-scores normalisation, in which the values are centred around the mean with a unit standard deviation.

Data normalisation was carried out to normalise the continuous variables (eg. age, heart rate, systolic blood pressure, diastolic blood pressure, total cholesterol, high-density lipoprotein, low-density lipoprotein, triglyceride, fasting blood glucose) by using z-score normalisation. Unscaled input variables may lead to a slow or unstable learning process. The purpose of data normalisation is to assure the consistency of continuous data, as well as to make database design easier and to ensure database security.

3.8 Data imputation

This study's definition of an incomplete dataset is a set of variables that is missing up to 50% of the time. The referenced missing dataset is for patient characteristics, not outcome data. The level of missingness in values across all variables was fully random and beyond control, because the dataset is a prospective dataset with retrospective data management. In this dataset, the probability of missingness is independent of neither the observed values in any variable nor the unobserved portion of the dataset.

As a result, the dataset is characterised as missing completely at random (MCAR), the maximum level of randomness, implying that the pattern of missing values is random and unrelated to any variable that may or may not be included in the analysis. Comprehensive

data were available for each of the outcomes. The testing dataset used for testing the model of the complete case was also used to test the imputed model.

Data imputation was carried out only for the best model of the ML and Dl models across three different time points. Secondary analyses were carried out after adding imputed missing cases. The imputation of the missing values was carried out using multivariate imputation by chained equations (MICE) with predictive mean matching (PMM) (Van Buuren & Groothuis-Oudshoorn, 2011). PMM is robust to assumptions breaches and backs up earlier recommendations for PMM and sample size (Van Buuren & Groothuis-Oudshoorn, 2011). The method works best with big samples and generates imputations that have many of the same properties as the original data. PMM is usually better at buffering any biases caused by distributional assumptions that have been violated (Kleinke, 2017). PMM can impute practically any type of variable, including continuous, semi-continuous, cluster, and panel data (Van Buuren & Groothuis-Oudshoorn, 2011; Vink et al., 2014). This method imputes missing values based on real values from other cases where predicted values are closest. This approach consists of a single variable, ***x***, which has missing data in some circumstances, and a group of variables, ***z*** (which do not contain missing data), which are used to impute ***x***, and the following steps were followed:

1) A linear regression of ***x*** on ***z*** was computed for cases with no missing data, yielding a set of coefficients ***b***.
2) The "posterior predictive distribution" of ***b*** was drawn at random, yielding a new set of coefficients ***b*****. A random draw from a multivariate normal distribution with mean b and the predicted covariance matrix of ***b*** (with an additional random draw for the residual variance) would be typical. This

phase is similar to all "proper" approaches for multiple imputations and was required to provide enough variety in the imputed values.

3) For all instances, both those with data missing on ***x*** and those with data present, the value of ***b**** was utilised to construct predicted values for ***x***.
4) A group of cases with observed ***x*** that have predicted values that are close to the anticipated value for the case with missing data was determined for each case with missing ***x***.
5) One value was chosen at random from among the close cases, and its observed value was assigned to replace the missing value.
6) For each full data set, steps 2 through 5 were repeated.

The objective of linear regression in PMM, unlike many other methods of imputation, is not to generate imputed values. Rather, it is used to create a statistic for comparing cases with missing data to cases with similar data. For this study, the number of cases was set to 3, that is, each case on ***x*** with missing data was matched to the three examples (with data) with the closest anticipated values. The ***x*** value of one of the three was assigned to the case with missing data at random. This method was carried out by R using the library (mice).

3.9 Baseline Characteristics

All analyses of baseline characteristics for STEMI and NSTEMI/UA patients were performed using the Statistical Package for Social Sciences (SPSS) version 16.0 programme. Continuous variables are represented in mean ± standard deviation while categorical variables are presented in frequency and percentage. To identify significant factors, a Chi-Square test (for categorical variables) and a two-sided independent student t-test (for continuous variables) were used in univariate analysis ($p < 0.001$).

3.10 Model development and hyperparameter tuning

In this study, 3 types of ML models were developed to predict the mortality of patients with STEMI or NSTEMI/UA which included LR, RF, and SVM algorithms. Each of the models was trained with all 50 (for STEMI) and 39 (for NSTEMI/UA) variables together with selected variables obtained through different feature selection techniques; SBE and RFE. This study applied the classification method which was used to predict the mortality of patients with STEMI and NSTEMI/UA accordingly with the output data of alive (1) and dead (0). As a result, 9 ML models were created for each STEMI and NSTEMI/UA dataset for the three time points (in-hospital, 30-days, and 1-year), as shown below:

Table 3.6: The list of ML models with their abbreviations.

ML models	Abbreviation
Logistic regression with all variables	LR
Logistic regression with stepwise regression analysis selected variables	LRstepwise-SBE-LR
Logistic regression with RFE selected variables	RFE-LR
Random forest with all variables	RF
Random forest with sequential backward elimination selected variables	RFvarImp-SBE-RF
Random forest with RFE selected variables	RFE-RF
Support vector machine with all variables	SVM

Table 3.6, continued.

ML models	Abbreviation
Support vector machine with sequential backward elimination selected variables	SVMvarImp-SBE-SVM
Support vector machine with RFE selected variables	RFE-SVM

Due to the rapid development in the usage of DL approaches in recent years, this study included DL as an addition to the work. It was also used to verify the viability of variables selected by the ML since feature learning is employed in DL instead of feature selection. Feature learning in DL can only learn all of the qualities that have been provided and accomplish the tasks that have been assigned to it, such as classification and detection. DL does not carry out the process of selecting only significant variables and reducing the size of the overall variables fed into it (Kwon et al., 2019a). For each of the dataset and time points, two DL models were built which is one with all the variables and the other one with the variables selected by the best ML models.

In this study, for each of the time points (in-hospital, 30 days, and 1-year), the best ML and DL models were compared against the TIMI risk score for the STEMI group and TIMI risk score for NSTEMI/UA group. The TIMI risk score for STEMI (Morrow et al., 2000) was developed using the Intravenous nPA for Treatment Infarcted Myocardium Early II trial to predict STEMI patient mortality at 30 days, while the TIMI risk score for NSTEMI/UA (Antman et al., 2000) was developed using the TIMI 11B clinical trial for the composite endpoint of mortality at 14 days. The predictive capacity of the TIMI risk score for STEMI was consistent throughout time, from 24 hours to one year following hospital

admission (Morrow et al., 2000). The TIMI risk scores were chosen for this study because of their ability to score ACS subtypes individually, whereas the GRACE score was developed to predict ACS mortality in general and was utilised to predict long-term (six months) mortality (Lee et al., 2021). Furthermore, the TIMI risk scores are the commonly used risk scores for mortality prediction of STEMI and NSTEMI/UA patients in Malaysian hospitals (MOH, 2011, 2019). Despite the fact that both TIMI risk scores were created in a clinical trial sample that was largely Western, only the TIMI risk score for STEMI has been validated against a Malaysian multi-ethnic group in the study by Selvarajah et al. (2012). As a result, the purpose of this study was to see how well ML and DL algorithms compare to the TIMI risk score in predicting the mortality of STEMI and NSTEMI/UA patients in Malaysia, as well as in the era when an early invasive revascularization strategy is becoming more popular.

The flow of developing predictive models utilising various DL and ML algorithms is depicted in the Figure 3.5 and Figure 3.4 respectively, below. It explains how models were developed, hyperparameters were modified, variables were selected, and the best model was chosen.

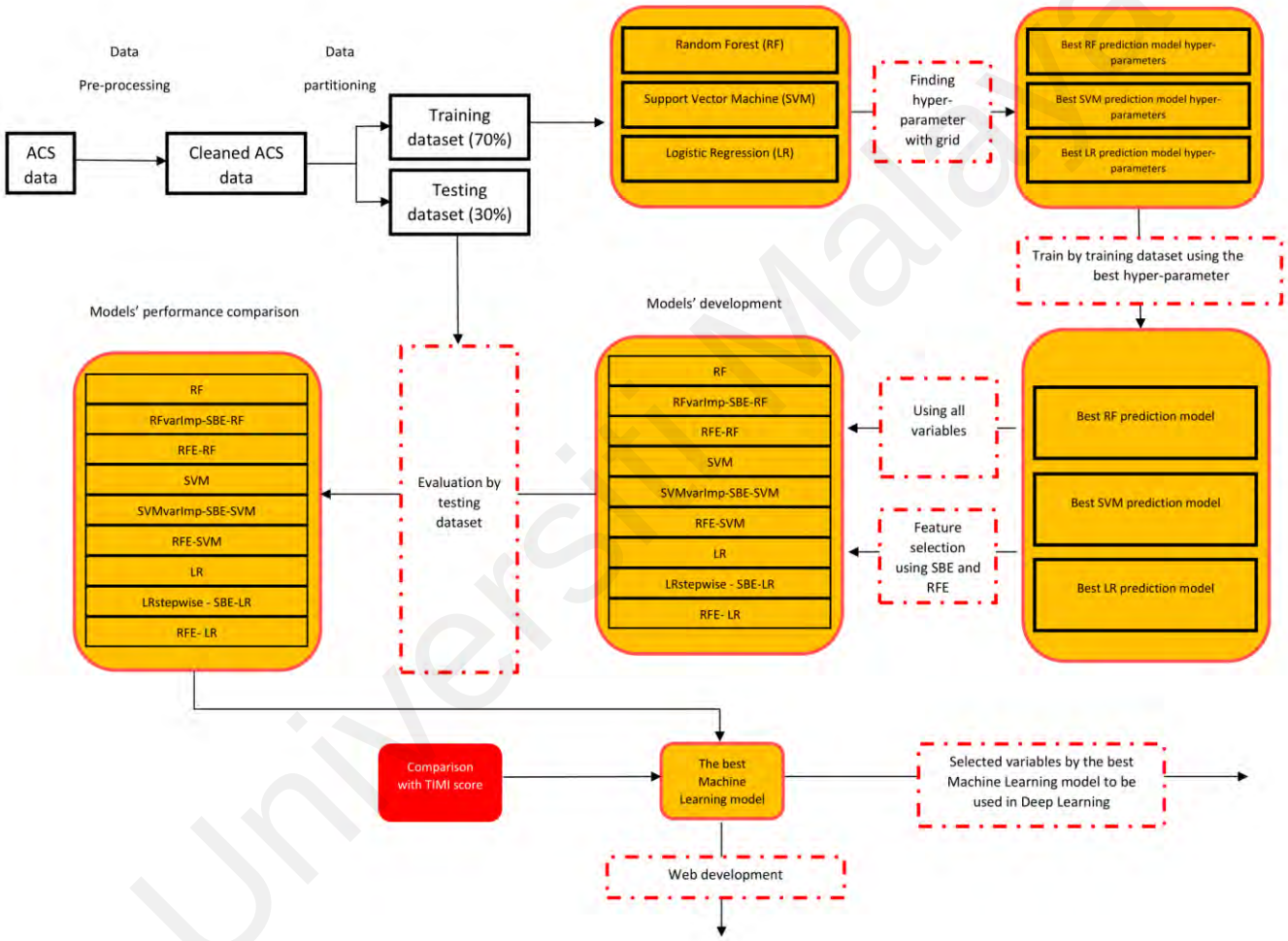


Figure 3.5: The flowchart of the ML predictive models' development.

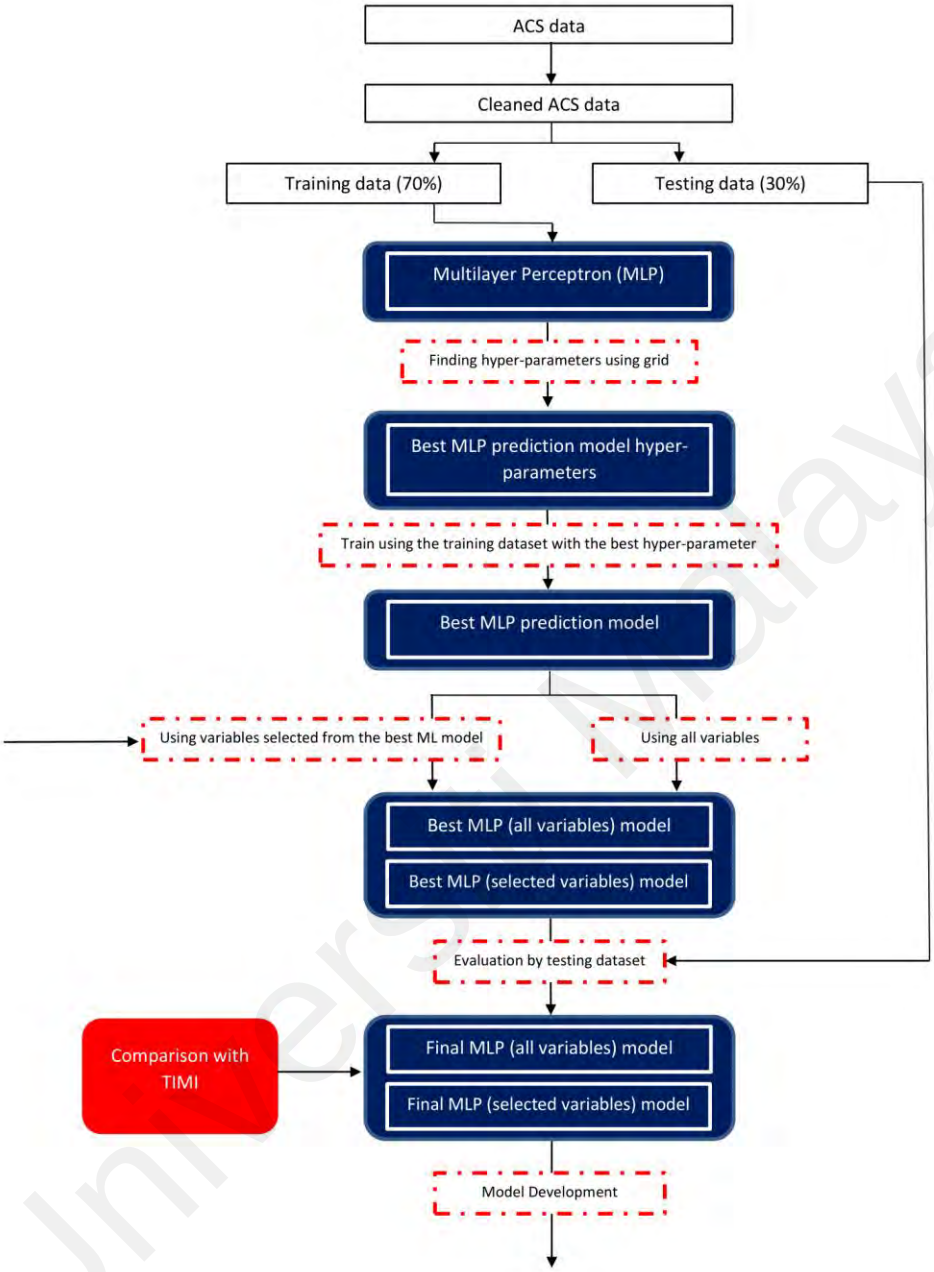


Figure 3.6: The flowchart of the DL predictive models' development.

3.10.1 Machine Learning Model Development

All of the ML models were constructed in R using the caret package's features. In this case, the resampling technique was applied to determine the optimum model by fine-tuning the parameter values. We used the following functions:

i) train_control (): To specify the resampling scheme

ii) train (): Act as the workhorse of caret, which takes several parameters.

(a) Logistic regression (LR)

In this study, the *glm ()* function was used to fit the logistic regression model, which is typically used to fit generalised linear models. However, in this situation, this study utilised this method to fit an LR model. Kassambara (2018) claims that *glm* has no tuning parameters. Hence, the default generalised linear model was used to conduct binary classifications on all 50 variables for STEMI and 39 variables for the NSTEMI/UA dataset, including those with and without features. This study used the family parameter, family = "*binomial*", to indicate that the output is a two-class categorical response. This instructs *glm ()* to fit an LR model rather than one of the many other models available to the *glm*. The *tunelength* function was set to 10 because a longer tune length allows the algorithm to examine more potential models and possibly find a better one. 10-fold cross validation was also carried out in the model building.

(b) Random Forest (RF)

RF is an ensemble method that uses classification or regression trees to make decisions. The function that was used for RF model building was *randomForest*. According to Kwon et al. (2019b), RF entails the creation of many decision trees. Each tree partitions the sample data by separating the variable at distinct cutting points. The data for the tree is chosen at random from the training dataset. It generates hundreds of different classification trees and then utilises them all together as a composite classifier. The majority rule is used to the votes of individual classifiers to determine the final classification of a given sample.

There are only two tuning parameters in RF: *ntree* and *mtry*. The number of trees to grow is specified by the parameter *ntree*, which must be large enough to provide OOB error stabilisation. The number of variables chosen randomly as candidates at each split is specified by the parameter *mtry*, which can range from 1 to the total number of variables. According to Chen et al. (2015), the prediction will not differ statistically if a high number of *ntree* are picked, but the model will take longer to build. In classification issues, the default value of *mtry* is $\sqrt{N}$, where N is the total number of variables. In general, using the default values for *ntree* and *mtry* can yield good results (Wang et al., 2021).

During the model development, the model was trained with a different number of *ntree* which comprises 500, 1000, 1500, and 2000 trees respectively. The optimal numbers of *mtry*, which were randomly picked at each split as the candidates, were determined using grid search with *tuneLength* =10 and 10-fold-cross validation. The model with the highest AUC was selected as the best. The algorithm of RF is shown below:

1) A bootstrap sample of the training set was used to grow each of the RF tree.
2) When building a tree, n number of variables were picked at random from N variables at each node.
3) It is proposed that the value of n be set to $n=\sqrt{N}$ and then increased until the OOB error is the smallest. For all values of n, one variable with the best split was used at each node.
4) RF method was then applied for testing data for prediction.
5) Steps 1 to 4 were then repeated using different values of *ntree* (500,1000,1500 and 2000) and *mtry* starting the default value of classification RF, $\sqrt{N}$ which is 7 for STEMI dataset and 6 for NSTEMI/UA dataset.
6) The best model was chosen as the one with the highest AUC.

(c) Support Vector Machine (SVM)

The SVM models were built using the *svmLinear*, *svmRadial*, and *svmPoly* techniques, respectively. Cost, C determines the possibility of misclassifications in the SVM model, therefore it is critical to implement a penalty for the model's inaccuracy. When the cost value is increased, the SVM model is less likely to misclassify a point.

The default SVM linear classifier used a cost value of C=1. The C value can be tuned accordingly. Optimum C value can be chosen via the highest ROC value. The degree in the polynomial kernel acts as the decision boundary of the hyperplane flexibility. Higher degree polynomial kernels allow a more flexible decision boundary. Whereas, the gamma, γ; value in both polynomial and radial kernel specifies how far a training sample's impact spreads. The inverse of the radius of the data impact chosen as support vectors by

the model is Gamma. This means that a high Gamma will only consider places close to the reasonable hyperplane, but a low Gamma will consider sites farther away.

The SVM model was built by utilising the default linear kernel function and SVM, as well as the radial and polynomial kernel functions. To discover the best parameter in each kernel function, grid search was employed with *tuneLength* =10, and 10-fold-cross validation was used. The best parameters were then used to train the whole training set and were applied for data testing. However, in this study, the SVM model that applied radial kernel functions demonstrated the highest value of AUC as compared to the other two kernel functions. Hence, the radial kernel function was adopted as the main kernel in building the SVM model for this study.

The optimised parameters for each of the ML models are shown in Table 3.7 and Table 3.8 below.

Table 3.7: The hyper-parameters values for optimum ML model performance for STEMI dataset.

In-hospital	
Machine Learning	**Parameters**
Random Forest	Seed = 42 mtry = default ($\sqrt{N} = 7$) Number of trees = 1000 Pre-processing = centre and scale Cross-validation=10 Number of iterations = 3
Support Vector Machine	Kernel = radial Seed = 42 Pre-processing = centre and scale

Table 3.7, continued.

In-hospital	
Machine Learning	**Parameters**
	C-value = 0.25 Gamma value = 0.01 Cross-validation=10
Logistic Regression	Seed = 42 Method= glm Cross-validation=10 Tune length = 10 Family = binomial
30-days	
Machine Learning	**Parameters**
Random Forest	Seed = 42 mtry = default ($\sqrt{N} = 7$) Number of trees = 1000 Pre-processing = centre and scale Cross-validation=10 Number of iterations = 3
Support Vector Machine	Kernel = radial Seed = 42 Pre-processing = centre and scale C-value = 0.25 Gamma value = 0.01 Cross-validation=10
Logistic Regression	Seed = 42 Method= glm Cross-validation=10

Table 3.7, continued.

30 days	
Machine Learning	**Parameters**
	Tune length = 10 Family = binomial
1-year	
Machine Learning	**Parameters**
Random Forest	Seed = 42 mtry = default ($\sqrt{N} = 7$) Number of trees = 1000 Pre-processing = centre and scale Cross-validation=10 Number of iterations = 3
Support Vector Machine	Kernel = radial Seed = 42 Pre-processing = centre and scale C-value = 0.25 Gamma value = 0.01 Cross-validation=10
Logistic Regression	Seed = 42 Method= glm Cross-validation=10 Tune length = 10 Family = binomial

Table 3.8: The hyper-parameters values for optimum ML model performance for NSTEMI/UA dataset.

In-hospital	
Machine Learning	**Parameters**
Random Forest	Seed = 42 mtry = default ($\sqrt{N} = 6$) Number of trees = 500 Pre-processing = centre and scale Cross-validation=10 Number of iterations = 3
Support Vector Machine	Kernel = radial Seed = 42 Pre-processing = centre and scale C-value = 0.20 Gamma value = 0.01 Cross-validation=10
Logistic Regression	Seed = 42 Method= glm Cross-validation=10 Tune length = 10 Family = binomial
30 days	
Machine Learning	**Parameters**
Random Forest	Seed = 42 mtry = default ($\sqrt{N} = 6$) Number of trees = 500 Pre-processing = centre and scale Cross-validation=10

Table 3.8, continued.

30 days	
Machine Learning	**Parameters**
	Number of iterations = 3
Support Vector Machine	Kernel = radial Seed = 42 Pre-processing = centre and scale C-value = 0.10 Gamma value = 0.02 Cross-validation=10
Logistic Regression	Seed = 42 Method= glm Cross-validation=10 Tune length = 10 Family = binomial
1-year	
Machine Learning	**Parameters**
Random Forest	Seed = 42 mtry = default ($\sqrt{N} = 6$) Number of trees = 1000 Pre-processing = centre and scale Cross-validation=10 Number of iterations = 3
Support Vector Machine	Kernel = radial Seed = 42 Pre-processing = centre and scale C-value = 0.30 Gamma value = 0.01

Table 3.8, continued.

1-year	
Machine Learning	**Parameters**
	Cross-validation=10
Logistic Regression	Seed = 42 Method= glm Cross-validation=10 Tune length = 10 Family = binomial

3.10.2 Feature Selection Methods

Maldonado and Weber (2009) claimed that feature selection is significant in classification issues since it reduces computing complexity and improves the generalisation of classifiers. The wrapper technique with SBE was utilised in their research because it adjusts better to a dataset and helps reduce overfitting. Any kernel functions can be utilised with sequential backward elimination in SVM.

Wrapper methods, according to Ang et al. (2015), will take relationships between features into account and deliver more accurate results than filter methods. According to Kumar and Minz (2014), the wrapper technique will utilise a learning algorithm to evaluate a subset. A subset will be chosen that is most suited to the learning algorithm. Balakrishnan et al. (2008) employed sequential backward elimination to find important variables for classifying diabetes and non-diabetic patients in their study. After the feature selection, the classification accuracy was found to be higher. Another commonly used wrapper method is the RFE method (Kohavi & John, 1997). According to Aziida et al. (2021), RFE performed better when combined with ML methods (such as RF and SVM) which yielded a high value

of AUC in predicting 30 days mortality of patients with STEMI. Similar results were observed from other studies on factors affecting amlodipine induced pedal edema and its classification (Chopra et al., 2017), selecting feature subsets in bioinformatics (Lin et al., 2018), feature selection workflow for high-dimensional omics data (Perez-Riverol et al., 2017) and identification of risk genes associated with myocardial infarction (Zhou et al., 2017). RFE is also computationally less complex than the other feature selection methods. The stepwise regression analysis is the common feature selection used for the LR algorithm. According to Juhan et al. (2019), stepwise regression analysis yielded a good feature selection to be employed in predicting risk factors for patients with STEMI (Juhan et al., 2019). This method is proven to be beneficial in selecting features that increase the model's performance by combining variables through backward or forward elimination from the variable importance ranking (Gibson et al., 2020; Loring et al., 2020; Tsai, 2009).

In this study, the wrapper method of SBE, RFE, and stepwise regression analysis were adopted to eliminate the irrelevant and insignificant features from the 50 variables of the STEMI dataset and 39 variables of the NSTEMI/UA dataset.

(a) Stepwise regression analysis

Stepwise regression is a methodology for fitting regression models in which the selection of predictive variables is automated. In this study, this feature selection method was utilised to develop the LR model by removing less important features and reducing the overall number of variables. It is an approach that allows changing the course by removing or adding variables at each stage. It has an option called direction, which can have the following values: "both", "forward", "backward", but in this study, the "backward" method was chosen to carry out the stepwise regression analysis in finding the best variables for LR model building.

Firstly, univariate LR analysis was performed on all variables in the training dataset during the model development phase. Variables were chosen for the backward stepwise regression analysis if they were statistically significant ($p<0.05$). Following that, AIC was used to undertake backward step-down selection.

The stepwise regression analysis was easily calculated using the MASS package's R function *stepAIC()*. This function performed model selection by AIC. AIC was utilised to simplify the model without affecting its performance much. *stepAIC* is a model selection tool that favours models with a lower AIC value.

The absolute value of the AIC has little importance. The AIC is similar to the adjusted R-squared in that, it penalises by removing more variables from the model. Hence, the AIC value of a subset of variables was examined to see if it was increasing or decreasing when more variables were removed. When there are numerous models to choose from, the one with the lowest AIC value is preferred. Thus, the *StepAIC* in the last step generated the best set of features with the lowest AIC value.

Lastly, the performance of the model with full variables and the model with stepwise logistic variables were compared. The best model is defined as the model that has the lowest classification error rate in predicting the class of new test data.

(b) Recursive Feature Elimination (RFE)

A recursive procedure that classifies features according to their relevance is known as RFE. RFE follows an automatic feature selection method. RFE is a feature evaluation approach that recursively eliminates low-importance features using the feature weights from the ML algorithms. In this study, the RF, SVM, and LR models were trained using the *rfFuncs* function in R with 10 folds cross-validation and 3 iterations for the RFE feature

selection method. RFE using the *rfFuncs* function performs based on Gini importance as the variable exclusion criterion. By splitting the variable under consideration at a certain node, Gini importance determines how well the samples can be assigned to the two output classes. The higher the value, the better the separation of instances into the two classes will be, and the more important the inspected features will be. The summarised steps for RFE method used in this study are as below:

1) The model was trained using the training data to obtain the feature's importance.
2) The model performance was evaluated.
3) The features were then ranked based on their Gini importance according to their classification contribution. This stage resulted in a ranking of variables.
4) The algorithm then formed a subset, *Si* of variables which keeps the most important variables. The subset, *Si* was then trained to obtain the new classification model performance.
5) Step 4 was repeated with a different set of variables by removing different variables iteratively until the subset, *Si* is empty.
6) The list of performance measurement values corresponding to each subset was produced and the subset with the highest accuracy was selected to be the best set of features.

As an explanation to the above steps, accuracy was determined using predictions derived using the trained algorithm's 10-fold cross-validation approach. This method gave a list of concatenated labels for predictions over 10 folds; for each in-fold instance, the prediction was made using the algorithm that was trained on all out-of-fold examples. The algorithm was retrained with new, reduced features and new partitioning after the feature with the lowest absolute weight was removed from the input dataset. It was then repeated until the

input dataset contained only one feature. After all iterations, the accuracy means, and standard deviation were computed as mentioned in step 5 above. The set of variables with the highest accuracy was then used for the model building of the classifiers RF, SVM, and LR.

(c) Sequential Backward Elimination (SBE)

The wrapper approach of SBE was used to remove irrelevant and insignificant characteristics from the 50 (for STEMI) and 39 (for NSTEMI/UA) variables in this study. This method was only carried out for the classifiers RF and SVM because the LR classifier backward elimination method was carried out using the stepwise regression analysis as discussed above. In order to eliminate the insignificant variables, all the variables were ranked according to their variable importance during the first training of the model by using the function *varImp()* from the caret package. The model was trained using 10-fold cross validation. Then, the variable importance graphs were plotted for better visualization. The backward elimination process was done by the following steps:

1) The initial model, *M1*, with all the variables, *N* in the training dataset was trained using the classifier and the performance measure (AUC) was recorded.
2) A graph of variable importance was then formed where the most important variable was ranked at the top and the least important variable at the bottom.
3) From the variable ranking graph of importance, the least important variable (the one at the bottom of the rank) was removed, which produced a new subset of variable, *N-1*.
4) The model with the remaining variables, *N-1* was trained again and the model's performance measure (AUC) was recorded.

5) Steps 3 and 4 were repeated by removing the remaining variables one by one from the bottom upwards to form new subsets, *N-2, N-3, N-4*, … *N-x*, until the most important variable in the dataset was left. The AUC values were recorded and compared after the training process of the subsets *N-2, N-3, N-4,... N-x*.
6) Significant variables from the backward elimination method were identified by plotting a line graph. A significant variable is defined as the variable that causes the drop of AUC value upon deletion, indicating that, by removing the variable, it decreases the model's performance, hence it is important for the variable to be in the model.
7) Next, a new cycle was started where a new model, *M2,* was formed consisting of only the significant variables.
8) Steps 1 to 6 were repeated for a few cycles to ensure that the new model maintained a higher AUC value than the previous model. This cycle ends when the new model's overall AUC value cannot be further improved by using the new set of variables.
9) The remaining variables, *N-x* in the last model, *Mx* (the model with the highest value of overall AUC) was selected to be incorporated as the variables for the final models of the respective models' development.

The figures below show the drop of AUC values when SBE was implemented to RF (Figure 3.7 for STEMI and Figure 3.9 for NSTEMI/UA) and SVM (Figure 3.8 for STEMI and Figure 3.10 for NSTEMI/UA) for all variables in this study. The x-axis indicates the list of variables ranked from the least important (left) to the most important (right) and the y-axis is the AUC value. The reason for this kind of visualisation of importance is to ease

the understanding and the detection of any drop of AUC value through graph line. The red line indicates the initial AUC value without the SBE being applied yet to the variables.

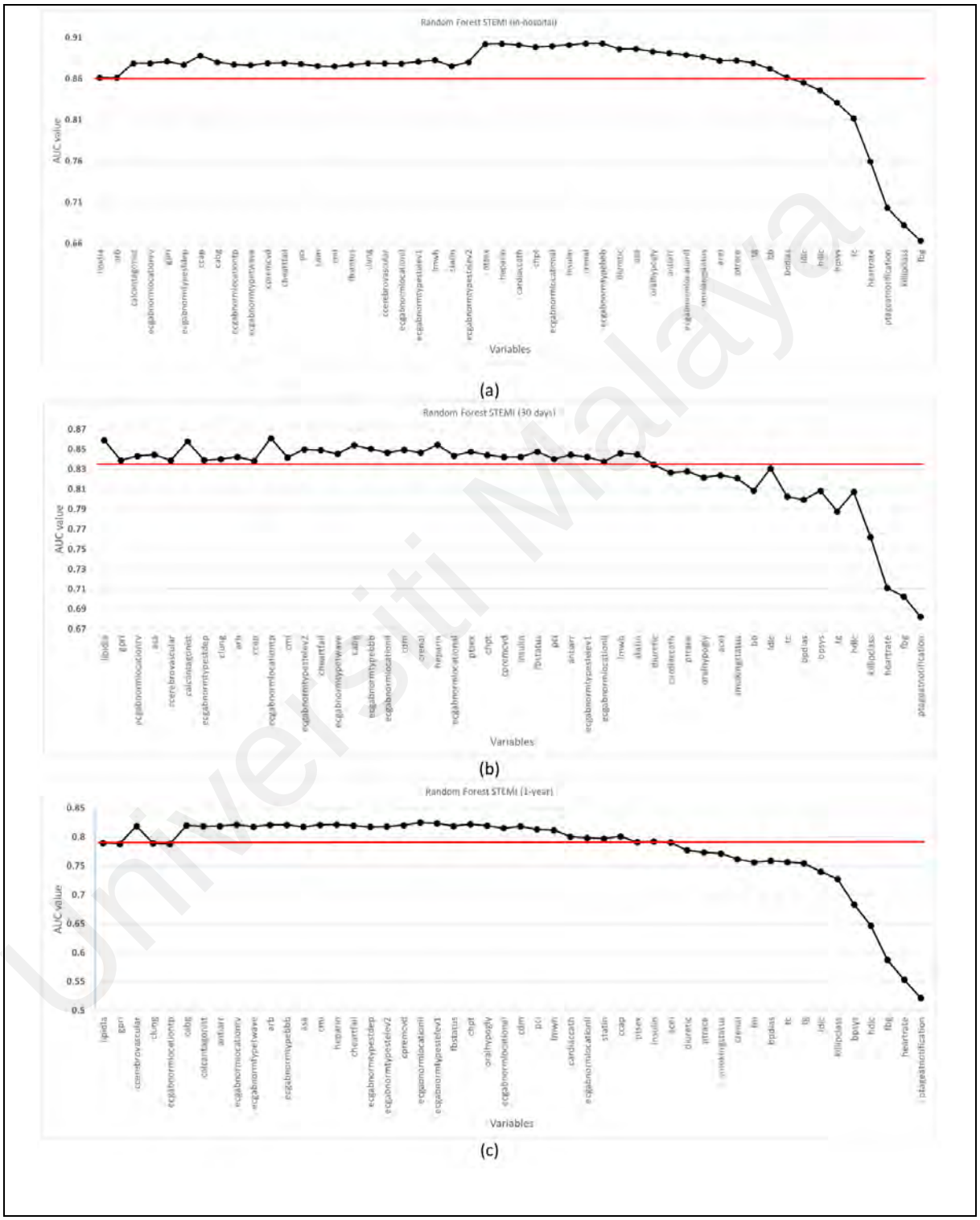


Figure 3.7: AUC values during the first cycle of feature selection for the RF model for STEMI dataset across three data frames (a) in-hospital, (b) 30 days, and (c) 1-year.

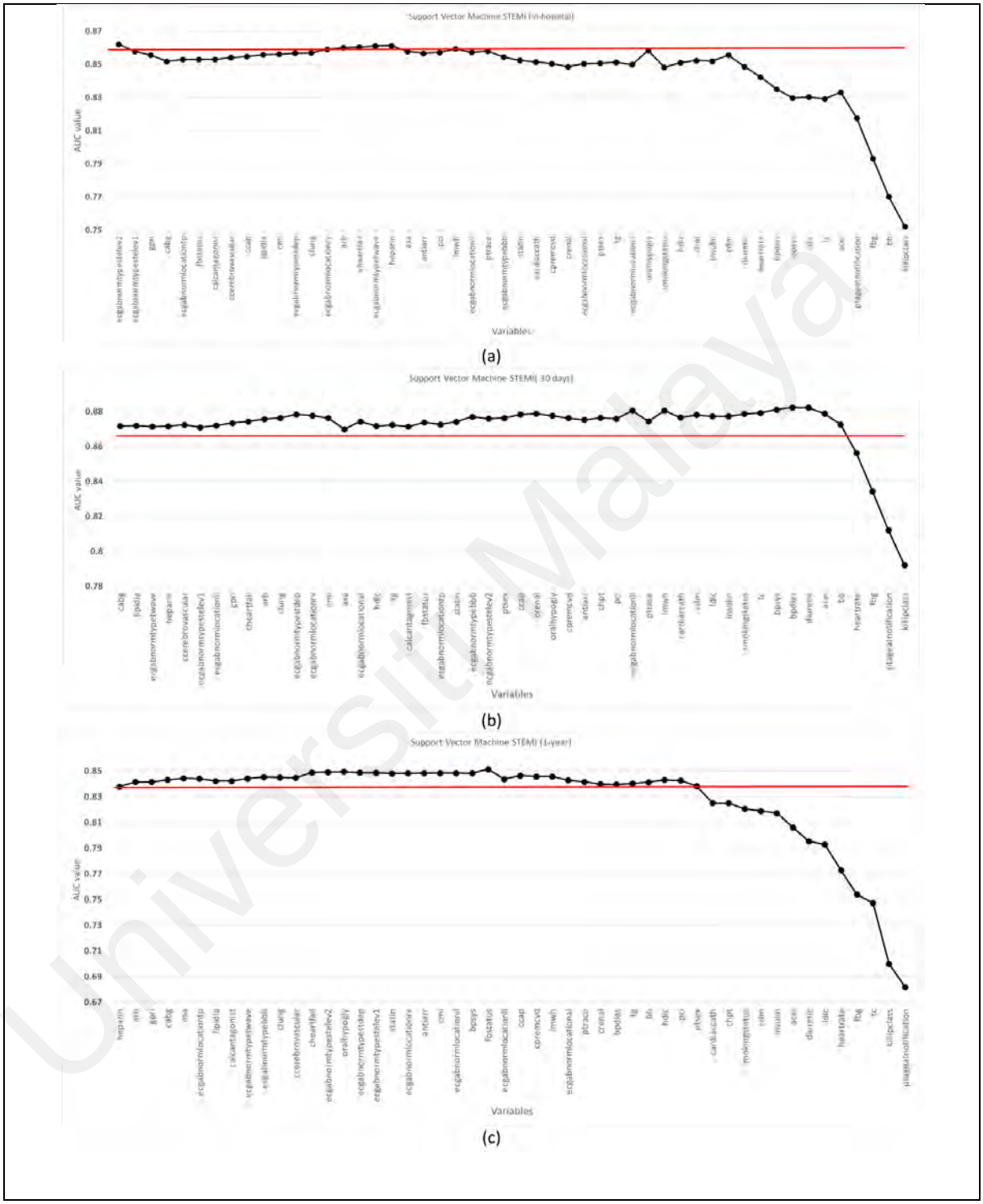


Figure 3.8: AUC values during the first cycle of feature selection for the SVM model for STEMI dataset across three data frames (a) in-hospital, (b) 30 days, and (c) 1-year.

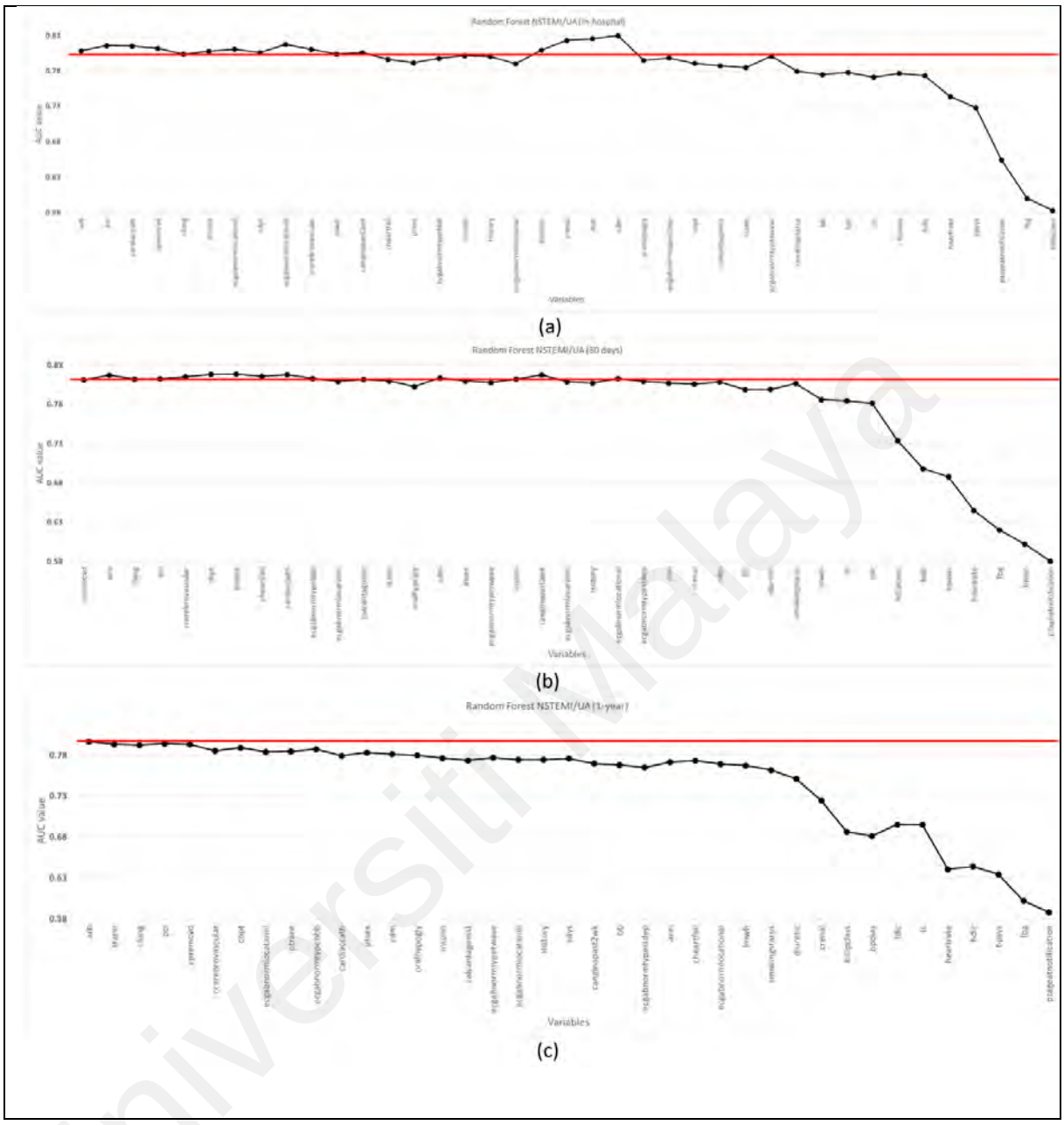


Figure 3.9: AUC values during the first cycle of feature selection for the RF model for NSTEMI/UA dataset across three data frames (a) in-hospital, (b) 30 days, and (c) 1-year.

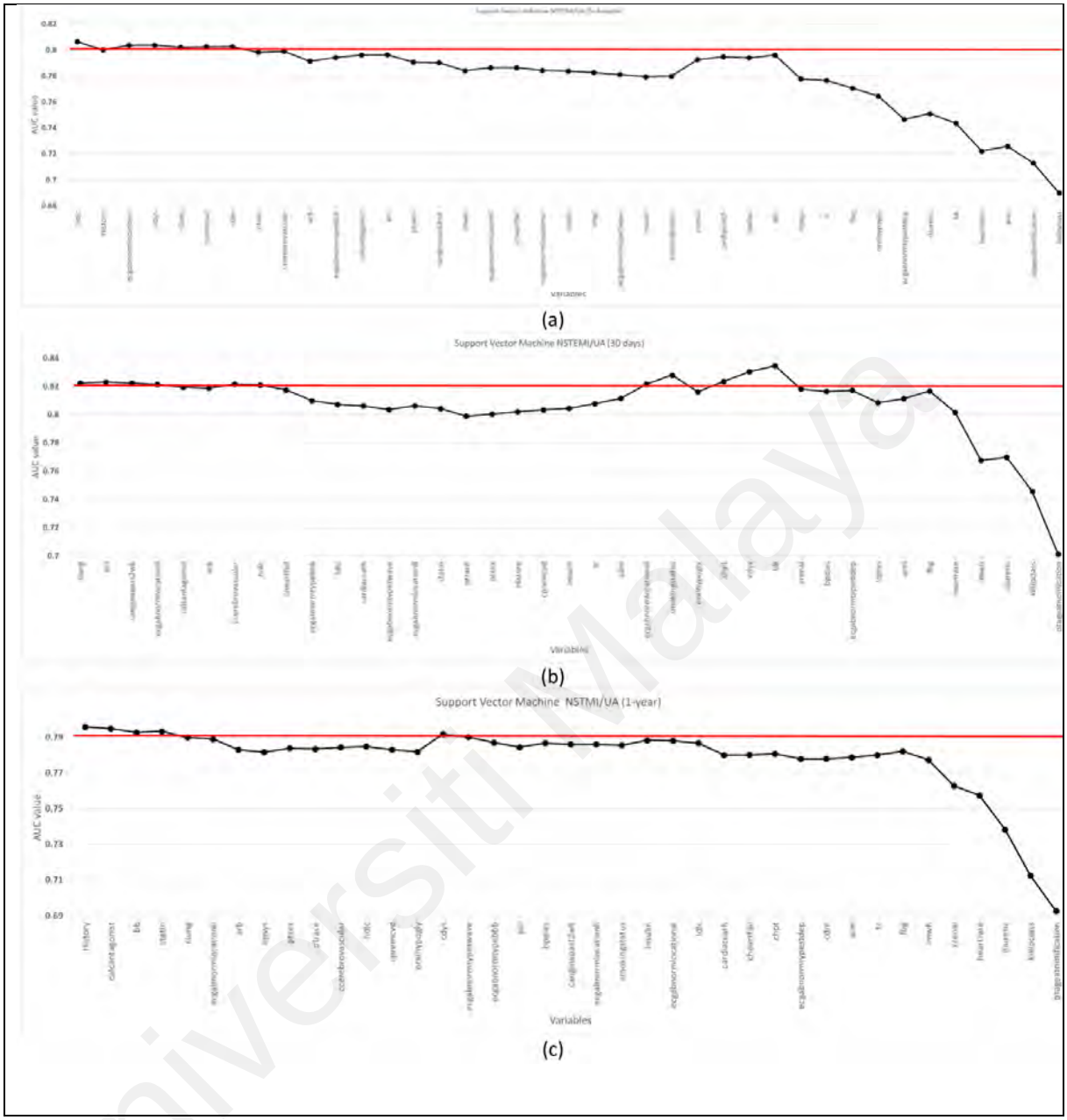


Figure 3.10: AUC values during the first cycle of feature selection for the SVM model for the NSTEMI/UA dataset across three data frames (a) in-hospital, (b) 30 days, and (c) 1-year.

As being explained above, this step was repeated until a newly forming group of variables' AUC is smaller than the previous forming group of variables. Once the AUC value of a newly forming group could not get any higher, the process was stopped and declared that the final forming group is the best model of the algorithm with the least number of variables. The ranking of the final best model of RF (Figure 3.11 for STEMI and

Figure 3.13 for NSTEMI/UA) and SVM (Figure 3.12 for STEMI and Figure 3.14 for NSTEMI/UA) are visualised below.

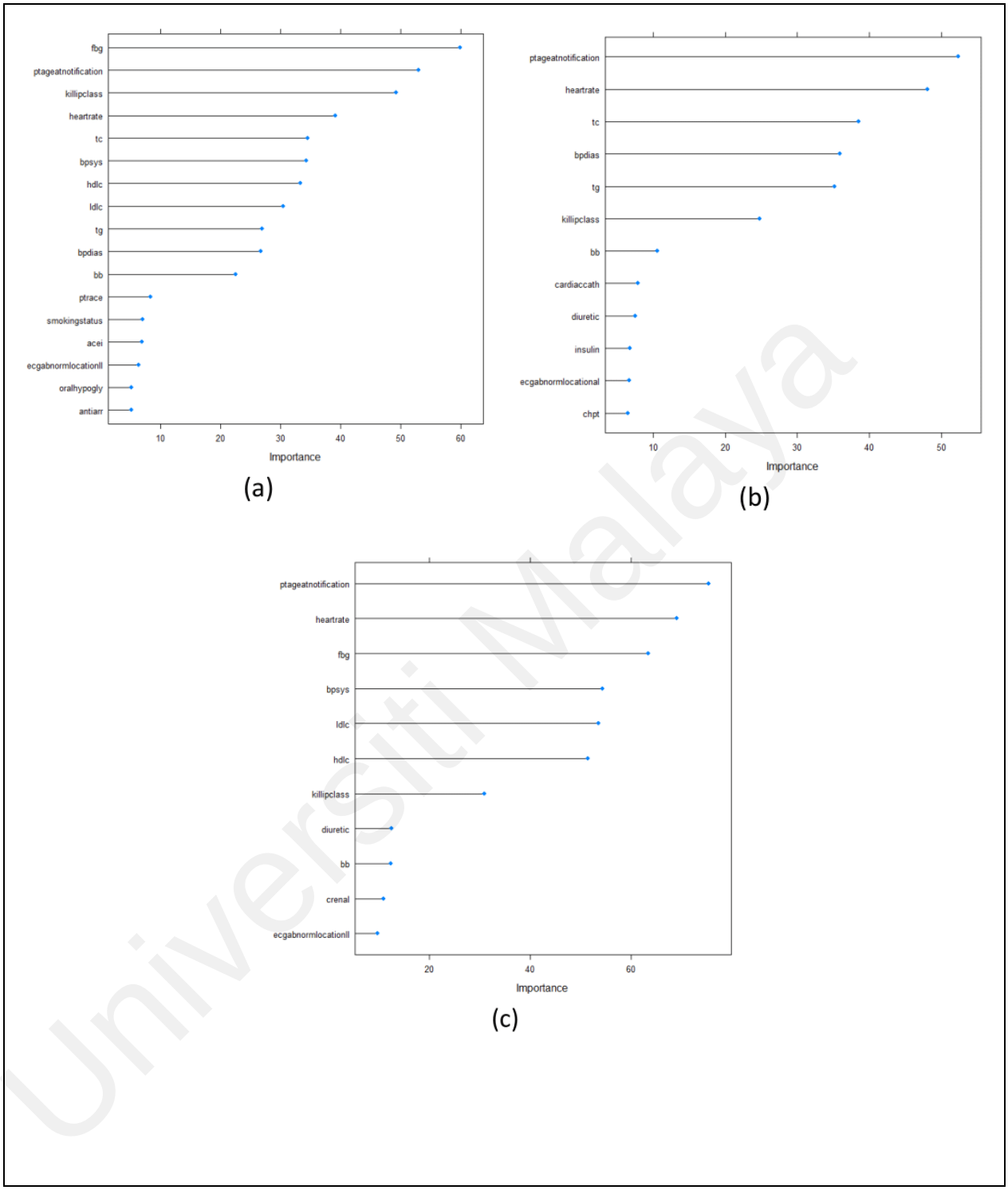


Figure 3.11: Ranking of the final models with selected variables for STEMI dataset in RF across three mortality prediction time frames (a) in-hospital, (b) 30 days, and (c) 1-year.

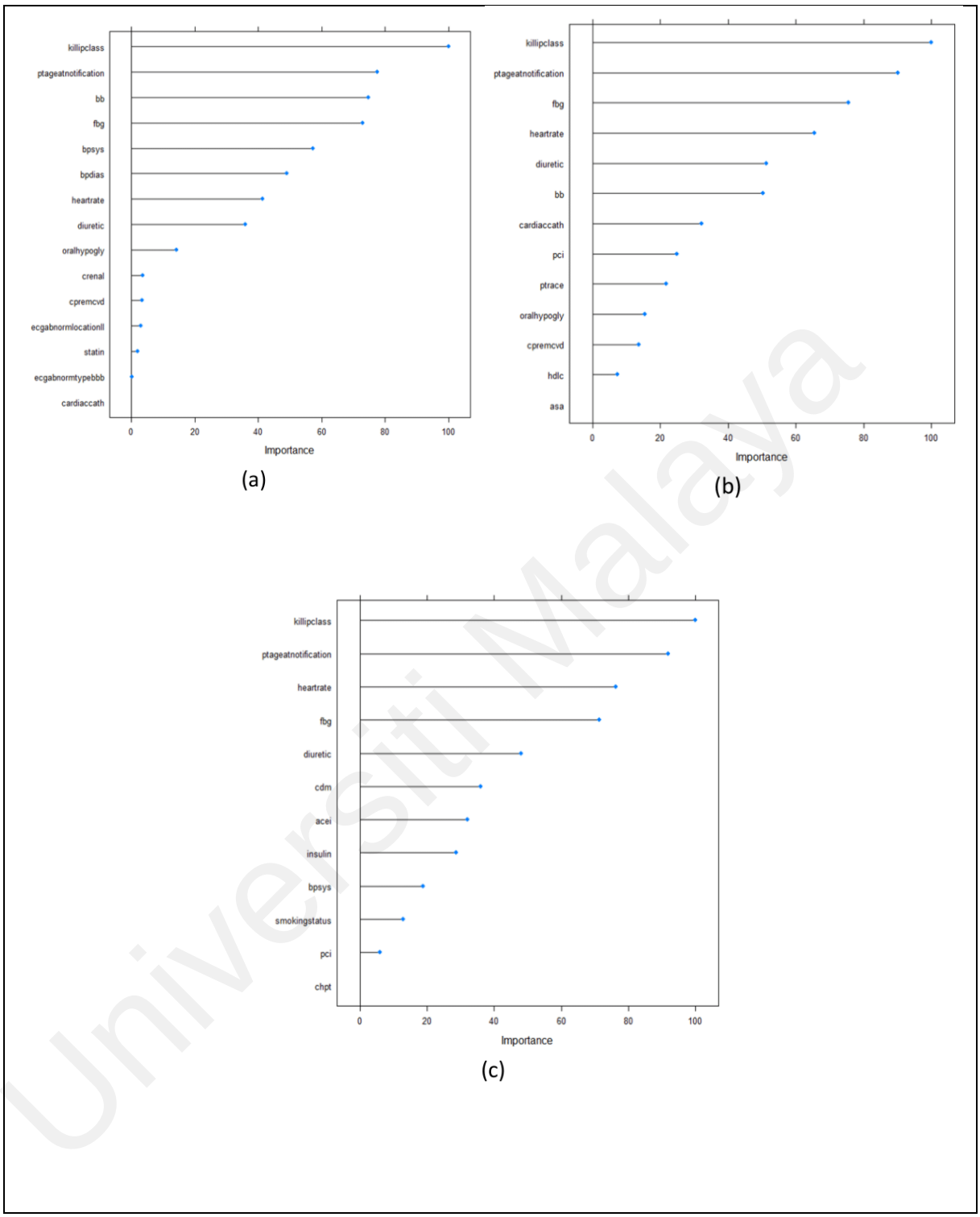


Figure 3.12: Ranking of the final models with selected variables for STEMI dataset in SVM across three mortality prediction time frames (a) in-hospital, (b) 30 days, and (c) 1-year.

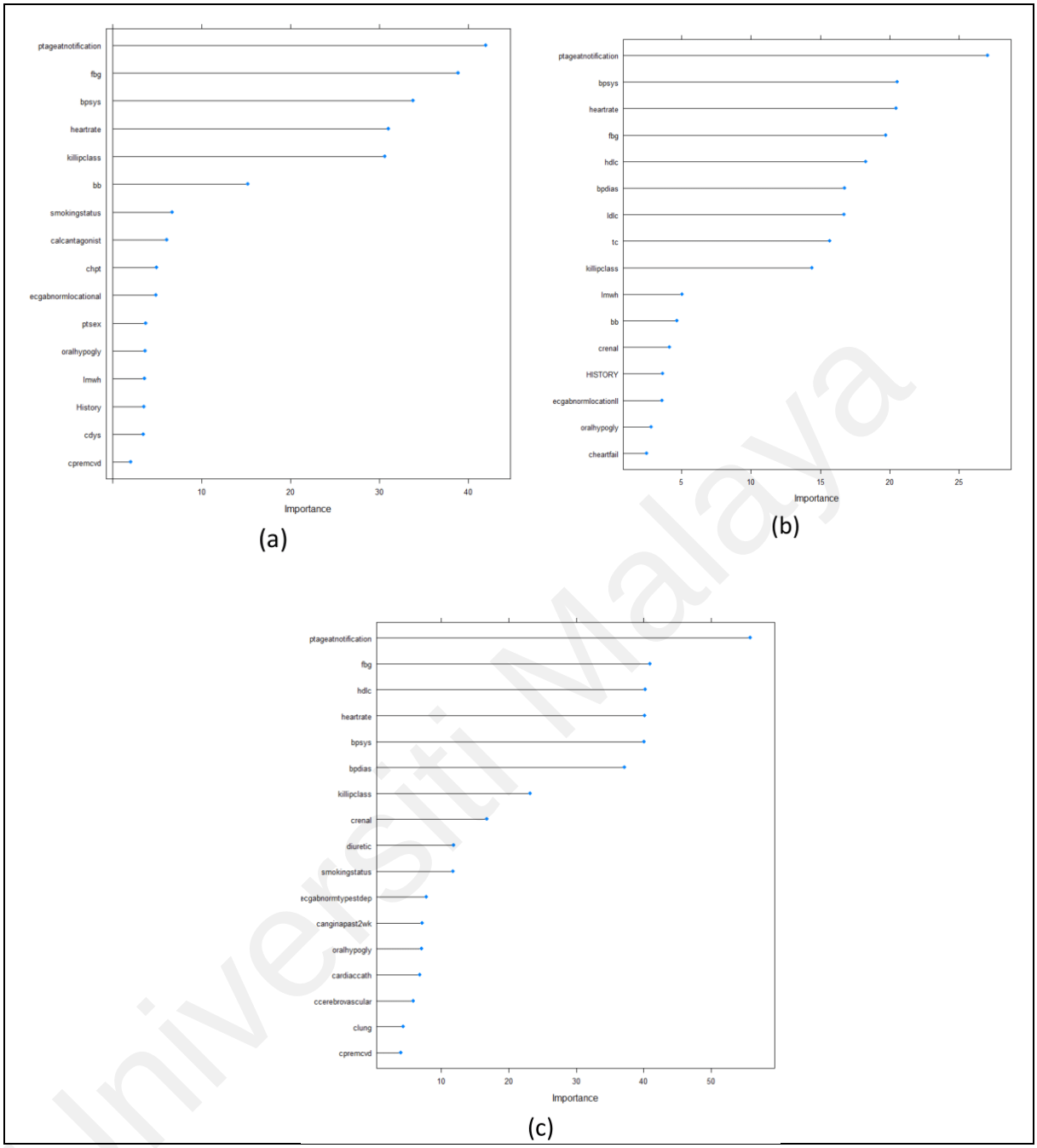


Figure 3.13: Ranking of the final models with selected variables for NSTEMI/UA dataset in RF across three mortality prediction time frames (a) in-hospital, (b) 30 days, and (c) 1-year.

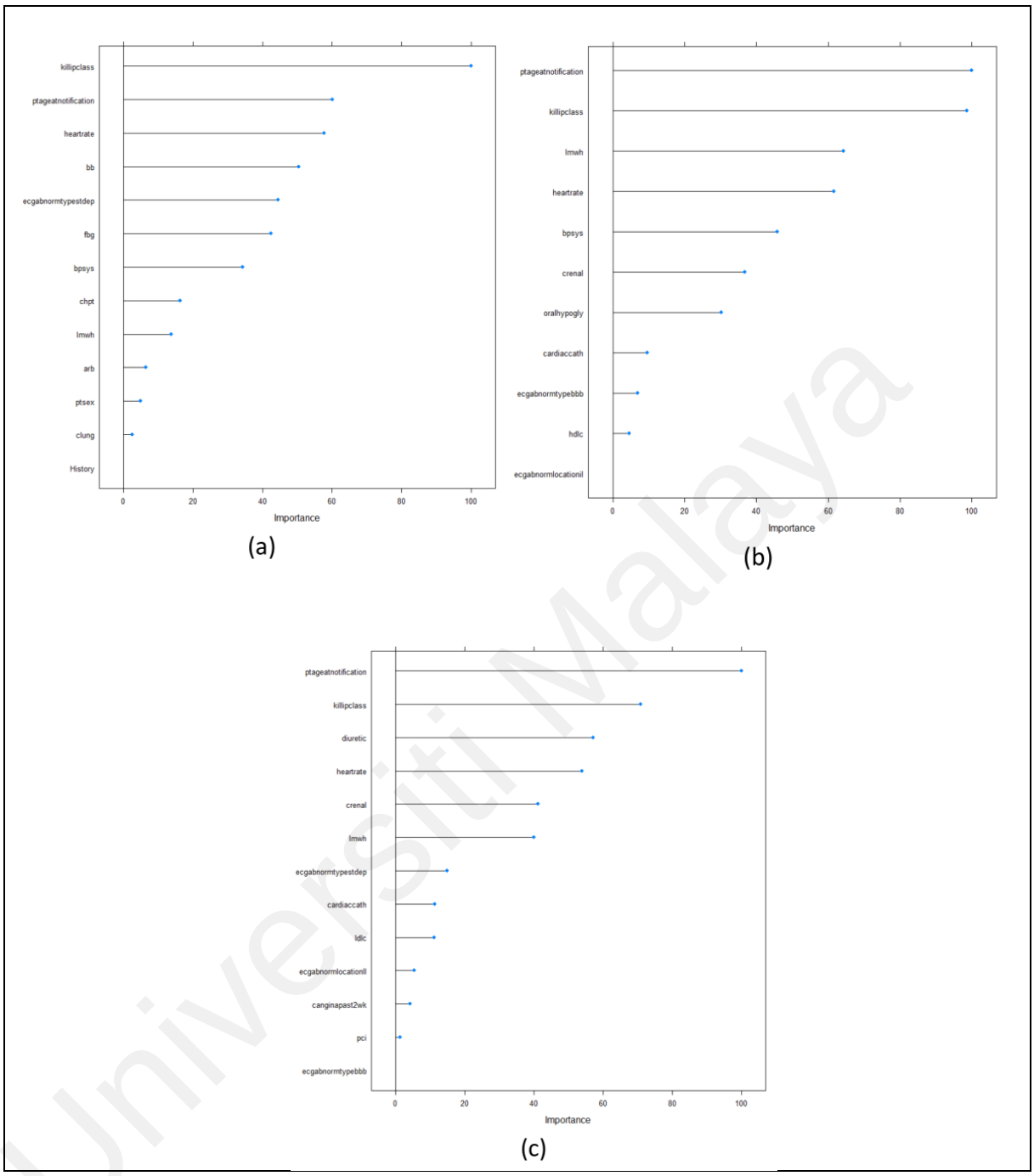


Figure 3.14: Ranking of the final models with selected variables for NSTEMI/UA dataset in SVM across three mortality prediction time frames (a) in-hospital, (b) 30 days, and (c) 1-year.

3.10.3 Deep Learning Model Development

3.10.3.1 Deep Learning Multilayer Perceptron Model Hyperparameter Tuning

As mentioned in the previous chapter, the DL model development of mortality prediction in this study is the extension to the future compass of the ML study as a whole. In this chapter, the applicability of DL in predicting mortality for patients with ACS will be discussed further in terms of algorithms and parameters in detail, as a broadening aspect of mortality prediction future enhancement.

According to Brownlee (2018), training a deep neural network model with backpropagation involves hyperparameter tuning. He also indicated that effective training of a model involved careful configuration, testing, and tuning the hyperparameters of the model and the learning process itself to best address the problem.

Hence, in the MLP model development, the hyperparameters such as numbers of nodes, numbers of hidden layers, dropout rate, learning rate, optimizer, batch size, epoch, and the activation function needed to be tuned carefully to maximise the performance of the predictive model. The possible tuning values/parameters that this study employed for each hyperparameter during the development of the MLP models are listed in Table 3.9 below.

Table 3.9: The possible values that were considered in the tuning process of the DL model development.

Hyper-Parameters	Possible Tuning Values
Number of nodes	(50, 135, 67, 33, 16, 2), (15, 64, 32, 16), (12, 32, 16, 8, 2)
Number of hidden layers	2,3,4
Dropout rate	0.1, 0.2, 0.3
Learning rate	0.1, 0.001, 0.0001
Optimizer	Adam, Adagrad
Batch size	8, 16, 32, 64

Table 3.9, continued.

Hyper-Parameters	Possible Tuning Values
Epoch	100,150,200,250,300
Activation function in hidden layers	ReLU, Tanh, linear, leaky ReLU
Activation function in output layer	sigmoid (binary classifications

For the DL models, hyperparameter tuning was done by manual tuning and grid search. The final DL model was selected based on several criteria:

1) The AUC and accuracy of the model

2) The sensitivity, specificity, and balanced accuracy of the model

3) The validation loss- training loss graph

4) The validation accuracy - training accuracy graph

5) The convergence of the two graphs

3.10.3.2 Deep Learning Model Building

The development of the DL model was carried out after the development of the ML models. ML model with the highest AUC value and the lowest number of selected variables was chosen as the best model. The rationale for developing the ML models first is to incorporate the best ML model's selected variables into the DL model development, as mentioned in the previous chapter by the study of Kwon et al., (2019a).

The table below summarises the hyperparameters used for the training of the DL models. The hyperparameters such as numbers of nodes, numbers of hidden layers, dropout rate, learning rate, optimizer, batch size, and epoch of each model were tuned so the graphs

of training loss and validation loss and the graph of training and validation accuracy converge.

For model construction, MLP based on DL incorporates four hidden layers, batch normalisation, and dropout layers. Three to four hidden layers (depending on the models in Table 3.10) were selected because adding more layers did not result in a substantial performance improvement. The number of nodes differs in each time point for both STEMI and NSTEMI/UA patients. This is because there is no general technique to determine the optimal number of neurons in DL. The neuron number was optimised by manual tuning and grid search.

For the DL models, Tensor Flow was utilised and Adam optimizer in R with default values and binary-cross entropy as the loss function (Kwon et al., 2019a). Adam optimiser was used instead of Adagrad optimiser because when Adagrad optimiser was used, the models become under-fitted with low specificity. Adam optimiser was simple to implement, and the default configuration parameters work well for the majority of problems. It has various features that combine the benefits of both Gradients with Momentum and RMSProp, such as low memory needs, suitability for non-stationary targets, and optimal performance with huge data and parameters.

The rectified linear unit (ReLU) as the activation function was used after comparing other activation functions, such as softmax, linear, Tanh, leaky ReLU, and exponential linear unit. The DL models used ReLU as the activation function on the hidden layers because others activation functions showed a decrease in the sensitivity of the model. ReLU is a type of activation function that is frequently employed in DL models. In other words, if the function receives a negative value, it returns 0; if it receives a positive value, it returns the same positive value. It also allows the DL model to account for non-linearities and

unique interaction effects. The ReLU function has the advantage of being a reasonably cheap function to compute due to its simplicity. The model can be trained and run in a short amount of time because there is no complicated arithmetic involved. Similarly, it converges faster, implying that the slope does not plateau as the input value increases.

The sigmoid function was used in the output layer and binary cross-entropy as the loss function because this study was performed on the binary classifications. The last layer (output) of a DL model can use a sigmoid function to turn the network's output into a probability score, which is easier to work with and analyse. Table 3.10 shows the final architecture of DL models for STEMI and NSTEMI/UA.

Table 3.10: The optimized hyper-parameters for DL model development.

STEMI						
	In-hospital		**30-days**		**1-year**	
	All	**Selected**	**All**	**Selected**	**All**	**Selected**
Number of hidden layers	4	3	4	3	4	4
Number of nodes in each layer	(50, 135, 67, 33, 16, 2)	(15, 64, 32, 16)	(50, 135, 67, 33, 16, 2)	(12, 32, 16, 8, 2)	(50, 135, 67, 33, 16, 2)	(12, 135, 67, 33, 16, 2)
Dropout rate	(0.5, 0.4, 0.3, 0.1 ,0.1)	(0.3, 0.3, 0.3, 0.3)	(0.5, 0.4, 0.3, 0.1 ,0.1)	(0.1, 0.4, 0.2, 0.3)	(0.5, 0.4, 0.3, 0.1 ,0.1)	(0.5, 0.4, 0.3, 0.1 ,0.1)
Learning rate	0.001	0.001	0.001	0.001	0.001	0.001
Optimiser	Adam					
Batch size	64	128	64	64	64	64
Epoch	100	80	100	80	100	100
Activation function	• Hidden layer: ReLU • Output layer: Sigmoid					
NSTEMI/UA						
	In-hospital		**30-days**		**1-year**	
	All	**Selected**	**All**	**Selected**	**All**	**Selected**

Table 3.10, continued.

NSTEMI/UA						
	In-hospital		**30-days**		**1-year**	
	All	**Selected**	**All**	**Selected**	**All**	**Selected**
Number of hidden layers	4	3	4	3	4	
Number of nodes in each layer	(39, 135, 67, 33, 16, 2)	(13, 64, 32, 16, 2)	(39, 135, 67, 33, 16, 2)	(11, 64, 32, 16, 2)	(39, 135, 67, 33, 16, 2)	(13, 64, 32, 16, 2)
Dropout rate	(0.5, 0.4, 0.3, 0.1 ,0.1)	(0.3, 0.4, 0.3, 0.5)	(0.2, 0.2, 0.4, 0.2, 0.4)	(0.3, 0.1, 0.3, 0.2)	(0.5, 0.4, 0.3, 0.1 ,0.1)	(0.5, 0.5, 0.4, 0.1)
Learning rate	0.001	0.001	0.001	0.001	0.001	0.001
Optimiser	Adam					
Batch size	64	32	64	64	64	32
Epoch	100	70	100	70	100	100
Activation function	• Hidden layer: ReLU • Output layer: Sigmoid					

The Figure 3.15 and Figure 3.16 below show the graph of training loss and validation loss in each epoch for the STEMI and NSTEMI/UA group respectively. The validation loss shows how well the model fits new data, while the training loss shows how well the model matches the training data. The total of errors for each sample in the training set is used to compute the training loss and the validation loss.

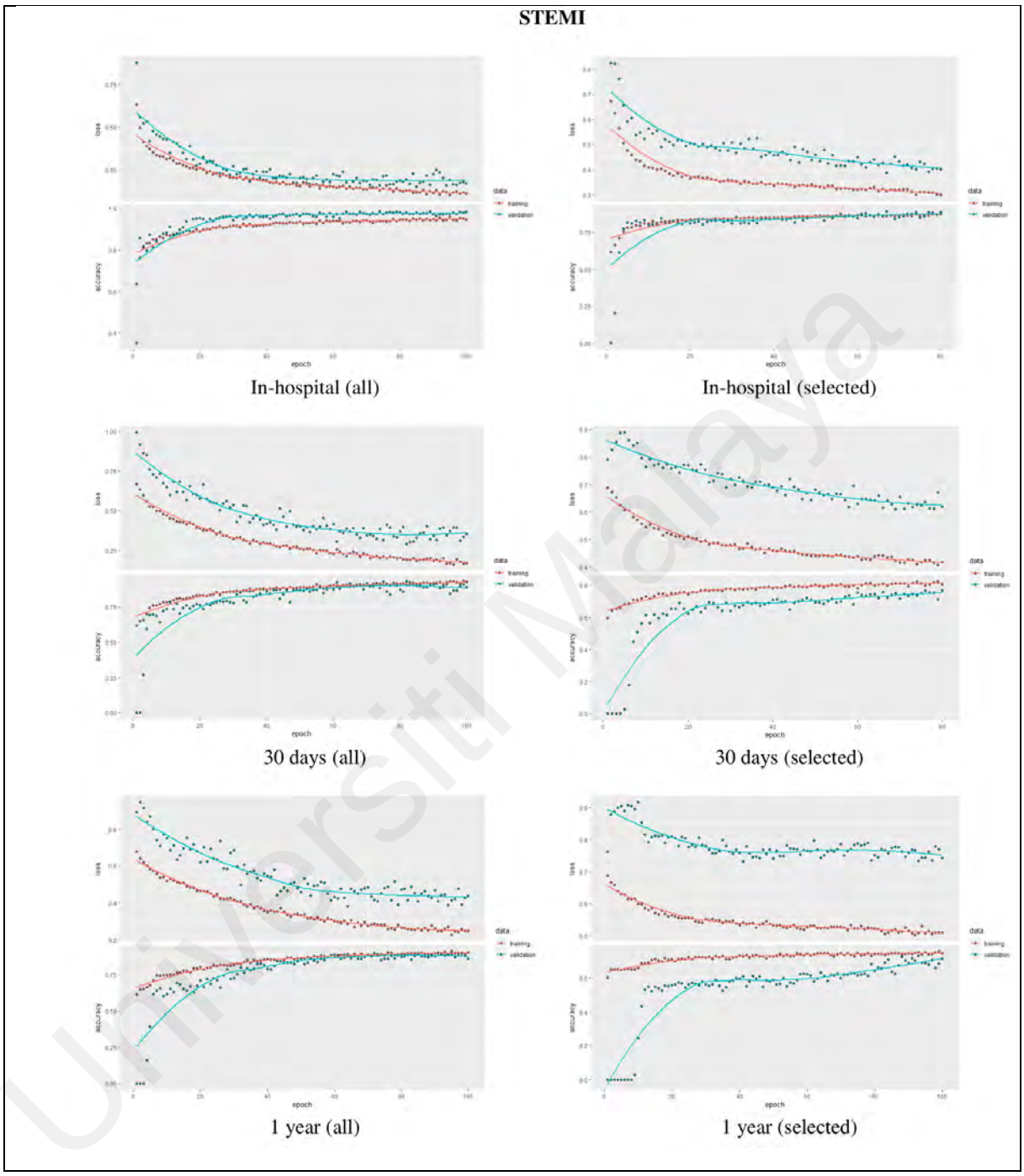


Figure 3.15: The graph of training loss and validation loss in each epoch for the STEMI group.

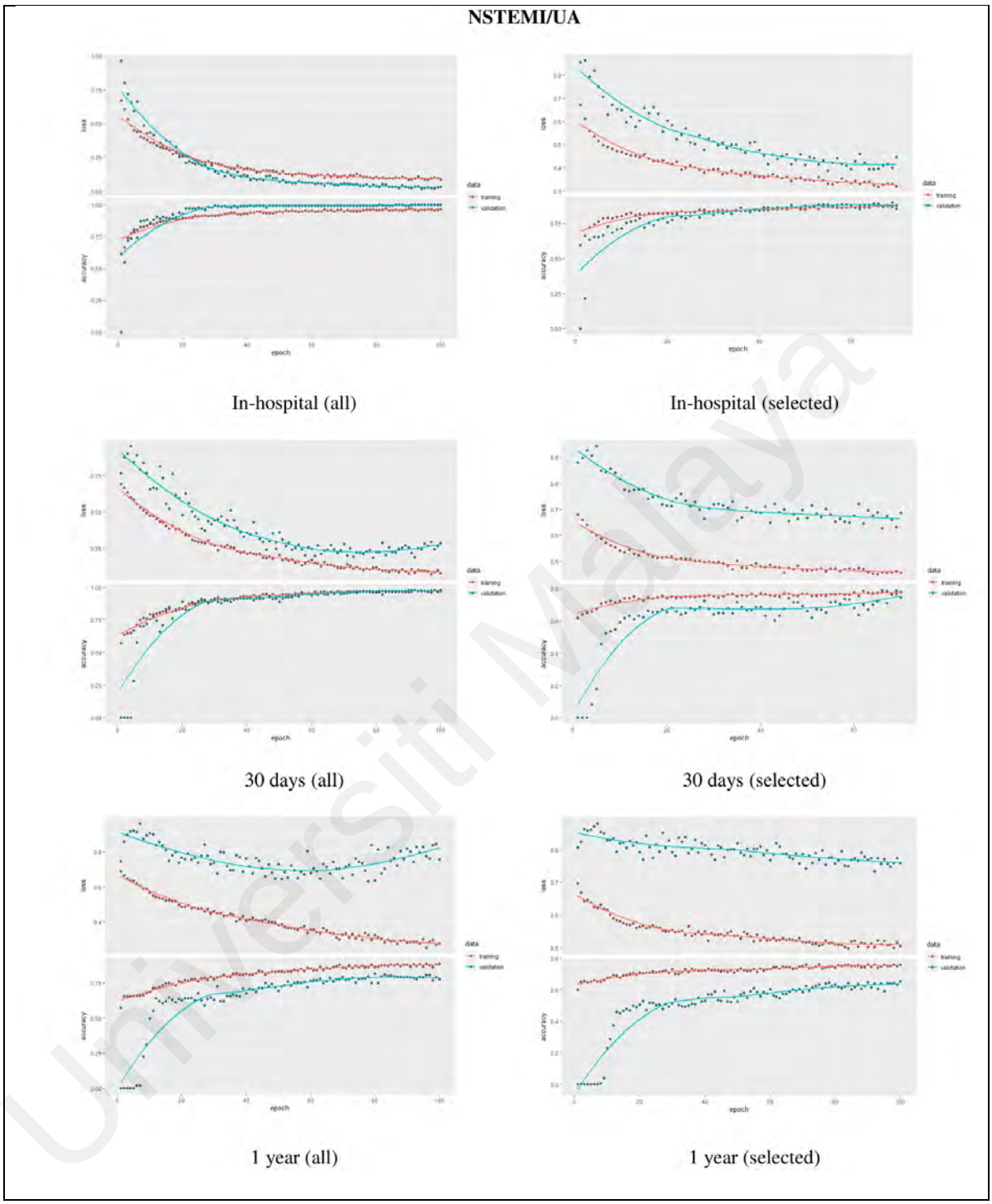


Figure 3.16: The graph of training loss and validation loss in each epoch for the NSTEMI/UA group.

3.11 Model Evaluation

3.11.1 Model calibration

The area under the Receiver Operating Characteristics (ROC) curve, or simply AUC, has been utilised in medical diagnosis since the 1970s, according to Huang and Ling (2005) and Fawcett (2006). They claimed that instead of accuracy, AUC should be used to evaluate and compare classifiers because AUC is a more accurate measure in general, and simple classification accuracy is usually a poor statistic for evaluating performance. This is because AUC will more directly and correctly reflect the ranking than accuracy. Values from both columns of the confusion matrix are used in metrics including accuracy, precision, lift, and *F* score. Even if the core classifier performance does not change, these measurements will alter when the class distribution changes. ROC graphs are not dependent on class distributions because they are based on TP and FP rates, with each dimension being a strict columnar ratio.

AUC was employed as an indicator in many of the medical diagnoses to assess the performance of the model they constructed. This was demonstrated in a study by Kwon et al. (2019a) in which AUC was employed as a comparison measure for mortality prediction in STEMI and NSTEMI patients. Wallert et al. (2017) employed AUC as a performance metric to predict two-year survivability following a first MI. Darabi et al. (2018), on the other hand, in their study, instead of using accuracy, employed AUC to anticipate the mortality risk of patients admitted to ICU. These studies chose AUC because it is unaffected by class imbalances.

The performance of the model was also determined using the confusion matrix. The list below shows the confusion matrix that was used to evaluate the performance of the

different classifiers. In this case of study, the positive class of the outcome was set to be 0 (dead). The confusion matrix gives information as below:

i) True Positives (TP): The dead patients predicted to die

ii) False Positives (FP): The alive patients predicted to die

iii) True Negatives (TN): The alive patients predicted to be alive

iv) False Negatives (FN): The dead patients predicted to be alive

Since AUC was used as the comparative metric in ACS mortality prediction, FNs are reasonably considered costlier than FPs. In the medical domain, it is preferable to predict a healthy person as sick, rather than a sick person as healthy, which causes the medical diagnosis to be failed. Hence, the precision of the positive class (0-died) should be emphasized more in the mortality prediction model. It is acceptable to scarify the precision of the negative class (1-survive).

Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and McNemar's test were taken into consideration when evaluating the model in this study. Among them, sensitivity and specificity define the true positive rate and true negative respectively. High sensitivity and specificity indicate that the correct classifications are made to correspond to each class.

3.11.2 Additional statistics

The results were expressed as mean and SD for continuous variables and as frequencies for categorical variables. Univariate analysis was performed using a Chi-Square test to identify significant variables and a two-sided independent student t-test ($p < 0.05$). The comparison between ML performances and with DL was done using a pair-wise corrected

resampled *t*-test (Dietterich, 1998; Raschka, 2018). Statistical significance was considered if the p-value was less than 0.05.

3.11.3 Comparison with TIMI score

Calculated TIMI scores were used from the NCVD registry for the validation data performance. TIMI score performance (in terms of AUC) was compared with the developed ML and DL models for all three time points using the 30% untouched testing dataset. ROC graphs were derived to compare performance between all ML models against TIMI risk score and the best ML model with the best DL model against TIMI risk score. Graphs on the rate of death according to the score (TIMI risk score) and percentiles (the best ML and DL models) were also derived to differentiate between the high- and low-risk patients based on cut-off points applicable in clinical practice and literature (Correia et al., 2014). A high risk of death was defined as a probability risk of death of more than 8% similar to reported by Correia et al. (2014). The rate of death graphs were also then tested for the trend in terms of a p-value.

3.11.4 Net reclassification improvement index (NRI)

After the cut-off value was found between the low and the high-risk patients on mortality for the best ML models and the best DL models using the percentage of death by Correia et al. (2014) study, NRI was then used. The cut-off values for the TIMI risk score for both STEMI and NSTEMI/UA were determined by the studies from Morrow et al. (2000) and Antman et al. (2000) respectively. The degree to which different mortality risk assessment approaches drive proper movement between categories was then evaluated by computing the NRI. An NRI can be interpreted as the percentage by which the net classification has improved by using a new different approach (Pencina et al., 2008). In this

study, NRI was used to determine the changes in discrimination between the TIMI risk score and ML/DL algorithm.

The NRI uses reclassification tables to examine whether there is an additive benefit gained from reclassifying patients using a different approach in mortality assessment. Since this study is looking at a binary target variable, the two-category NRI was utilised to evaluate the best models using equation 2.14 where the 'event' indicates when a patient is dead after a given time point (either in-hospital, 30 days, or 1-year), 'non-event' means the patient is still alive. The symbol 'up' implies that the old model (TIMI risk score for STEMI or NSTEMI/UA) classified the patient as low risk, whereas the new model (the best ML model or the best DL model) classified the patient as high risk. Whereas 'down' indicates the old model classified the patient as high risk while the new model classified as low risk.

An NRI of zero implies that the new model's discriminatory ability was equal to that of the previous model. A negative NRI indicates the new model was not able to discriminate between low-risk and high-risk as well as the old model, while a positive NRI suggests that the new model discriminated more accurately. The NRI is a number that runs from -2 to 2. An NRI of 2 implies that the old model successfully identified no patients, whereas the new model accurately differentiated between all high-risk and low-risk patients. An NRI of -2 shows that the new model correctly categorised none of the patients, while the previous model correctly identified all high- and low-risk patients. The p-value was then calculated using ANOVA between the probability of the ML and DL models mortality prediction and the TIMI risk score.

CHAPTER 4: RESULTS

This chapter presents all the findings and results that were obtained throughout the study which included the patients' characteristics for both STEMI and NSTEMI/UA dataset, performance results of the ML and DL models training with all and selected variables, feature selection of the models, performances of the TIMI risk scores for both STEMI and NSTEMI/UA as compared to the trained ML/DL model and application of the algorithms in clinical practice.

4.1 Patient characteristics

A total of 33,378 STEMI cases were identified, with 12,368 being complete cases (6,299 cases for in-hospital, 3,130 cases for 30 days, and 2,939 cases for 1 year). Table 4.1 illustrates the patients' characteristics used in this study on the complete dataset. The mean age was 56 years old. The majority of patients (~87%) were males. The overall mortality reported for in-hospital, 30 days, and 1-year was 5.4%, 8.1%, and 14.4% respectively. There was a high significant difference between survivors to non-survivors for in-hospital, 30-days and 1-year mortality in terms of gender, smoking status, diabetes, renal disease, heart rate, Killip class, fasting blood glucose, ECG abnormalities, beta-blockers, ACE inhibitors, statin, diuretics, insulin and anti-arrhythmic agent use ($p<0.0001$ for all). The patients' characteristics for secondary analysis using an imputed dataset for the STEMI dataset are shown in Appendix B. Both statistical analyses on the complete and imputed dataset were almost similar.

Table 4.1: Baseline characteristics for in-hospital, 30-days, and 1-year of the STEMI dataset.

Variables	Description	In-hospital				30 days				1-year			
		Total	Survival	Non-survival	*p*-value	Total	Survival	Non-survival	*p*-value	Total	Survival	Non-survival	*p*-value
N		6299	5961 (94.6)	338 (5.4)		3130	2878 (91.9)	252 (8.1)		2939	2516 (85.6)	423 (14.4)	
Age		55.8 ± 11.5	55.4 ± 11.3	63.8 ± 12.0	0.81	56.6 ± 11.7	56.0 ±11.4	64.2 ±12.5	0.054	56.6 ± 11.6	55.5 ± 11.2	62.8 ± 12.0	**0.028**
Race	Malay	3574 (56.7)	3365 (56.5)	209 (61.8)	**0.050**	1763 (56.3)	1608 (55.9)	155 (61.5)	**0.003**	1625 (55.3)	1370 (54.5)	255 (60.3)	**0.004**
	Chinese	1194 (19.0)	1126 (18.9)	68 (20.1)		552 (13.6)	498 (17.3)	54 (21.4)		531 (18.1)	453 (18.0)	78 (18.4)	
	Indian	1217 (19.3)	1170 (19.6)	47 (13.9)		640 (20.5)	602 (20.9)	38 (15.1)		610 (20.8)	530 (21.1)	80 (18.9)	
	Others	314 (5.0)	300 (5.0)	14 (4.1)		175 (5.6)	170 (5.9)	5 (2.0)		173 (5.9)	163 (6.5)	10 (2.4)	
Gender	Male	5417 (86.0)	5152 (86.4)	265 (78.4)	**<0.0001**	2681 (85.7)	2486 (86.4)	195 (77.4)	**<0.0001**	2533 (86.2)	2214 (88.0)	319 (75.4)	**<0.0001**
	Female	882 (14.0)	809 (13.6)	73 (21.6)		448 (14.4)	392 (13.6)	57 (22.6)		406 (13.8)	302 (12.0)	104 (24.6)	
Smoking status	Never	2003 (31.8)	1866 (31.3)	137 (40.5)	**<0.0001**	1053 (33.6)	941 (32.7)	112 (44.4)	**<0.0001**	977 (33.2)	786 (31.2)	191 (45.2)	**<0.0001**
	Former (quit tobacco > 30days)	1019 (16.2)	952 (16.0)	67 (19.8)		472 (15.1)	424 (14.7)	48 (19.0)		440 (15.0)	371 (14.7)	69 (16.3)	
	Current (tobacco < 30 days)	3277 (52.0)	3143 (52.7)	134 (39.6)		1605 (51.3)	1513 (52.6)	92 (36.5)		1522 (51.8)	1359 (54.0)	163 (38.5)	
History of hypertension		3344 (53.1)	3112 (52.2)	232 (68.6)	**<0.0001**	1697 (54.2)	1538 (53.4)	159 (63.1)	**0.003**	1587 (54.0)	1316 (52.3)	271 (64.1)	**<0.0001**
History of diabetes		2482 (39.4)	2291 (38.4)	191 (56.5)	**<0.0001**	1271 (40.6)	1129 (39.2)	142 (56.3)	**<0.0001**	1187 (40.4)	945 (37.6)	242 (57.2)	**<0.0001**
Family history of premature cardiovascular disease		892 (14.2)	869 (14.6)	23 (6.8)	**<0.0001**	435 (13.9)	419 (14.6)	16 (6.3)	**<0.0001**	410 (14.0)	372 (14.8)	38 (9.0)	**0.0001**
History of myocardial infarction		625 (9.9)	580 (9.7)	45 (13.3)	**0.032**	299 (9.6)	271 (9.4)	28 (11.1)	0.380	278 (9.5)	231 (9.2)	47 (11.1)	0.210
Documented CAD		583 (9.3)	552 (9.3)	31 (9.2)	0.956	358 (11.4)	323 (11.2)	35 (13.9)	0.202	341 (11.6)	273 (10.9)	68 (16.1)	**0.002**
History of heart failure		124 (2.0)	109 (1.8)	15 (4.4)	**0.001**	56 (1.8)	49 (1.7)	7 (2.8)	0.217	49 (1.7)	32 (1.3)	17 (4.0)	**<0.0001**
Chronic lung disease		114 (1.8)	101 (1.7)	13 (3.8)	**0.004**	61 (1.9)	54 (1.9)	7 (2.8)	0.321	60 (2.0)	44 (1.7)	16 (3.8)	**0.006**
Chronic renal disease		191 (3.0)	158 (2.7)	33 (9.8)	**<0.0001**	104 (3.3)	77 (2.7)	27 (10.7)	**<0.0001**	98 (3.3)	52 (2.1)	46 (10.9)	**<0.0001**
Cerebrovascular disease		171 (2.7)	156 (2.6)	15 (3.3)	**0.045**	88 (2.8)	80 (2.8)	8 (3.2)	0.716	84 (2.9)	63 (2.5)	21 (5.0)	**0.005**

Table 4.1, continued.

Variables	Description	In-hospital				30 days				1-year			
		Total	Survival	Non-survival	*p*-value	Total	Survival	Non-survival	*p*-value	Total	Survival	Non-survival	*p*-value
Heart rate		82.4 ± 21.1	81.7 ± 20.6	93.9 ± 26.6	**<0.0001**	82.9 ± 20.9	81.9 ± 20.0	94.5 ±27.0	**<0.0001**	82.6 ±20.6	81.1 ± 19.6	91.7 ± 24.2	**<0.0001**
Systolic blood pressure		132.8 ± 27.8	135.6 ± 27.4	120.4 ± 30.2	**0.011**	134.9 ± 28.2	135.6 ±28.0	126.4 ± 29.4	**<0.0001**	153.1 ± 28.0	135.9 ± 27.3	130.1 ± 3.1	**0.010**
Diastolic blood pressure		82.8 ± 94.1	83.4 ± 96.5	73.6 ± 20.2	0.965	81.3 ± 18.5	81.8 ±18.3	76.2 ±19.8	**<0.0001**	81.5 ± 18.4	82.1 ±18.1	78.4 ±20.0	0.066
Killip class	I	4300 (68.3)	4210 (70.6)	90 (26.6)	**<0.0001**	2072 (66.2)	1998 (69.4)	74 (29.4)	**<0.0001**	1980 (67.4)	1809 (71.9)	141 (40.4)	**<0.0001**
	II	1190 (18.9)	1132 (19.0)	58 (17.2)		558 (17.8)	506 (17.6)	52 (20.6)		512 (17.4)	413 (16.4)	99 (23.4)	
	III	237 (3.8)	200 (3.4)	37 (10.9)		128 (4.1)	98 (3.4)	30 (11.9)		110 (3.7)	71 (2.8)	39 (9.2)	
	IV	572 (9.1)	419 (7.0)	153 (45.3)		372 (11.9)	276 (9.6)	96 (38.1)		337 (11.5)	223 (8.9)	114 (27.0)	
Total cholesterol		5.4 ± 1.6	5.4 ± 1.6	4.8 ± 1.7	0.10	5.2 ± 1.4	5.3 ± 1.3	4.9 ± 1.6	**<0.0001**	5.2 ± 1.4	5.3 ± 1.3	4.9 ± 1.6	**0.005**
HDL		1.1±1.2	1.1 ± 1.2	1.0 ±0.3	0.952	1.1±0.4	1.1±0.4	1.1±0.3	0.140	1.1 ± 0.3	1.1 ±0.3	1.1 ±0.4	0.130
LDL		3.8 ± 10.7	3.8 ± 11.0	3.0 ± 1.4	0.706	3.5 ± 5.4	3.5 ± 5.6	3.1 ±1.4	0.295	3.1 ± 1.2	3.4 ± 1.2	3.1 ±1.4	**0.026**
Triglycerides		1.8 ± 1.7	1.8 ± 1.7	1.7 ± 1.0	0.529	1.7 ± 0.9	1.7 ± 0.9	1.7 ±0.9	0.762	1.7 ± 0.9	1.7 ±0.9	1.6 ±0.8	0.206
Fasting blood glucose		8.7 ± 4.4	8.5 ± 4.1	12.3 ± 6.7	**<0.0001**	8.8 ± 4.4	8.6 ± 4.1	11.7 ±6.5	**<0.0001**	8.8 ± 4.2	8.4 ± 3.8	10.8 ±5.8	**<0.0001**
ECG abnormalities type	ST-segment elevation ≥1mm in ≥2 contiguous limb leads	2804 (44.5)	2565 (44.6)	148 (43.8)	0.782	1518 (48.5)	1403 (48.7)	115 (45.6)	0.343	1437 (48.9)	1241 (49.3)	196 (46.3)	0.255
	ST-segment elevation ≥2mm in ≥2 contiguous frontal leads or chest leads	3373 (59.9)	3563 (59.8)	210 (62.1)	0.389	1828 (58.4)	1664 (57.8)	164 (65.1)	**0.025**	1710 (58.2)	1447 (57.5)	263 (62.2)	0.072
	ST-segment depression ≥0.5mm in ≥2 contiguous leads	627 (10.0)	589 (9.9)	38 (11.2)	0.416	280 (8.9)	254 (8.8)	26 (10.3)	0.426	267 (9.1)	219 (8.7)	48 (11.3)	0.080
	T-wave inversion ≥1mm	394 (6.3)	378 (6.3)	16 (4.7)	0.235	197 (6.3)	184 (6.4)	13 (5.2)	0.439	189 (6.4)	152 (6.0)	37 (8.7)	**0.036**
	Bundle branch block	138 (2.2)	111 (1.9)	27 (8.0)	**<0.0001**	72 (2.4)	56 (1.9)	18 (7.1)	**<0.0001**	59 (2.0)	38 (1.5)	21 (5.0)	**<0.0001**
ECG abnormalities location	Inferior leads: II, III, aVF	2998 (47.6)	2859 (48.0)	139 (41.1)	**0.014**	1520 (48.6)	1415 (49.2)	105 (41.7)	**0.022**	1433 (48.8)	1251 (49.7)	182 (43.0)	**0.011**
	Anterior leads: V1 to V4	3435 (54.5)	3233 (54.2)	202 (59.8)	**0.047**	1655 (52.9)	1498 (52.1)	157 (62.3)	**0.022**	1545 (52.6)	1287 (51.2)	258 (61.0)	**<0.0001**

Table 4.1, continued.

Variables	Description	In-hospital				30 days				1-year			
		Total	Survival	Non-survival	*p*-value	Total	Survival	Non-survival	*p*-value	Total	Survival	Non-survival	*p*-value
	Lateral leads: I, aVL, V5 to V6	1396 (22.2)	1295 (21.7)	101 (29.9)	**<0.0001**	744 (23.8)	659 (22.9)	85 (33.7)	**<0.0001**	705 (24.0)	567 (22.5)	138 (32.6)	**<0.0001**
	True posterior: V1, V2	515 (8.2)	484 (8.1)	321 (9.2)	0.492	258 (8.2)	235 (8.2)	23 (9.1)	0.595	243 (8.3)	213 (8.5)	30 (7.1)	0.343
	Right ventricle: ST elevation in lead V4R	524 (8.3)	494 (8.3)	30 (8.9)	0.703	286 (9.1)	262 (9.1)	24 (9.5)	0.824	269 (9.2)	231 (9.2)	38 (9.0)	0.896
FB status		4530 (71.9)	4288 (71.9)	242 (71.6)	0.893	2144 (68.5)	1973 (68.6)	171 (67.9)	0.819	1989 (67.7)	1717 (68.2)	272 (64.3)	0.109
Cardiac catheterization		2950 (46.8)	2812 (47.2)	138 (40.8)	**0.045**	1727 (55.2)	1619 (56.3)	108 (42.9)	**<0.0001**	1629 (55.4)	1455 (57.8)	174 (41.1)	**<0.0001**
PCI		2414 (38.3)	2298 (38.6)	116 (34.3)	0.120	1396 (44.6)	1304 (45.3)	92 (36.5)	**0.007**	1323 (45.0)	1188 (47.2)	135 (31.9)	**<0.0001**
CABG		33 (0.5)	30 (0.5)	3 (0.9)	0.341	33 (1.1)	27 (0.9)	6 (2.4)	**0.032**	22 (0.7)	18 (0.7)	4 (0.9)	0.611
ASA		6180 (98.1)	5862 (98.3)	318 (94.1)	**<0.0001**	3070 (98.1)	2829 (98.3)	241 (95.6)	**0.003**	2883 (98.1)	2473 (98.3)	410 (96.9)	0.058
GP receptor inhibitor		173 (2.7)	162 (2.7)	11 (3.3)	0.557	62 (2.0)	56 (1.9)	6 (2.4)	0.635	58 (2.0)	51 (2.0)	7 (1.7)	0.611
Heparin		962 (15.3)	900 (15.1)	62 (18.3)	0.107	549 (17.5)	501 (17.4)	48 (19.0)	0.512	523 (17.9)	459 (18.2)	64 (15.1)	0.121
LMWH		1546 (24.5)	1450 (24.3)	96 (28.4)	**0.09**	479 (15.3)	414 (14.4)	65 (25.8)	**<0.0001**	406 (13.8)	313 (12.4)	93 (22.0)	**<0.0001**
Beta blockers		4066 (64.5)	3978 (66.7)	88 (26.0)	**<0.0001**	1896 (60.6)	1800 (62.5)	96 (38.1)	**<0.0001**	1754 (59.7)	1558 (61.9)	196 (46.3)	**<0.0001**
ACE inhibitors		3320 (52.7)	3251 (54.5)	69 (20.4)	**<0.0001**	1509 (48.2)	1452 (50.5)	57 (22.6)	**<0.0001**	1388 (47.2)	1268 (50.4)	120 (28.4)	**<0.0001**
Angiotensin II receptor blocker		181 (2.9)	176 (3.0)	5 (1.5)	0.115	61 (1.9)	55 (1.9)	6 (2.4)	0.605	52 (1.8)	43 (1.7)	9 (2.1)	0.564
Statin		6013 (95.5)	5713 (95.8)	300 (88.8)	**<0.0001**	3003 (95.9)	2774 (96.4)	229 (90.9)	**<0.0001**	2820 (96.0)	2433 (96.7)	387 (91.5)	**<0.0001**
Other lipid lowering agent		127 (2.0)	124 (2.1)	3 (0.9)	0.129	51 (1.6)	48 (1.7)	3 (1.2)	0.566	47 (1.6)	38 (1.5)	9 (2.1)	0.349
Diuretics		1349 (21.4)	1201 (20.1)	148 (43.8)	**<0.0001**	720 (23.0)	610 (21.2)	110 (43.7)	**<0.0001**	651 (22.2)	473 (18.8)	178 (42.1)	**<0.0001**
Calcium antagonist		367 (5.8)	352 (5.9)	15 (4.4)	0.263	183 (5.8)	176 (6.1)	7 (2.8)	**0.030**	161 (5.5)	139 (5.5)	22 (5.2)	0.787
Oral hypoglycaemic agent		1345 (21.4)	1312 (22.0)	33 (9.8)	**<0.0001**	597 (19.1)	567 (19.7)	30 (11.9)	**0.003**	546 (18.6)	478 (19.0)	68 (16.1)	0.153

Table 4.1, continued.

Variables	Description	In-hospital				30 days				1-year			
		Total	Survival	Non-survival	*p*-value	Total	Survival	Non-survival	*p*-value	Total	Survival	Non-survival	*p*-value
Insulin		1658 (26.3)	1516 (25.4)	142 (42.0)	**<0.0001**	869 (27.8)	757 (26.3)	112 (44.4)	**<0.0001**	804 (27.3)	624 (24.8)	180 (42.6)	**<0.0001**
Anti-arrhythmic agent		313 (5.0)	276 (4.6)	37 (0.9)	**<0.0001**	178 (5.7)	144 (5.0)	34 (13.5)	**<0.0001**	151 (5.1)	114 (4.5)	37 (8.7)	**<0.0001**

The p-value is statistically highly significant as $p < 0.001$

In other respects, the total number of NSTEMI/UA patients found in this study was 42,683, with only 9,477 of these being complete cases. From the total amount of complete cases, 4,771 cases were in-hospital, 2,402 were 30 days and 2,304 were 1-year. The patients' characteristics were tabulated in Table 4.2. From Table 4.2, it was identified that the mean age of the NSTEMI/UA patients across the three-time points was 61 years old and the majority of them were males (~71%). The overall mortality rate reported for in-hospital was 3.65%, 6.83% for 30 days, and 17.23% for the 1-year dataset. Survivors significantly differed from non-survivors at all three time points in age, history of chronic renal disease, heart rate, Killip class, fasting blood glucose, ECG ST-segment depression, ACE inhibitors intake, and diuretics intake ($p < 0.0001$). The data imputation method was also applied to the NSTEMI/UA dataset and the summary characteristics can also be referred to in Appendix C. Both the complete dataset and the imputed dataset reported nearly similar statistical analyses. Subchapter 4.2 below discusses a further analysis of the imputed data.

Table 4.2: Baseline characteristics for in-hospital, 30-days, and 1-year of the NSTEMI/UA dataset.

Variables	Description	In-hospital				30 days				1-year			
		Total	Survival	Non-survival	*p*-value	Total	Survival	Non-survival	*p*-value	Total	Survival	Non-survival	*p*-value
N		4771	4597 (96.35)	174 (3.65)		2402	2238 (93.17)	164 (6.83)		2304	1907 (82.77)	397 (17.23)	
Age		61.30±11.84	61.05±11.79	67.80±11.42	**<0.0001**	61.69±12.05	61.13±11.85	69.35±12.17	**<0.0001**	61.66±12.08	60.31±11.79	68.15±11.34	**<0.0001**
Race	Malay	2063 (43.2)	1985 (43.2)	78 (44.8)		1039 (43.3)	962 (43.0)	77 (47.0)		992 (43.1)	808 (42.4)	184 (46.3)	
	Chinese	1356 (28.4)	1290 (28.1)	66 (37.9)		673 (28.0)	618 (27.6)	55 (33.5)		642 (27.9)	530 (27.8)	112 (28.2)	
	Indian	1047 (21.9)	1025 (22.3)	22 (12.6)		494 (20.6)	470 (21.0)	24 (4.9)		483 (21.0)	406 (21.3)	77 (19.4)	
	Others	305 (6.4)	297 (6.5)	8 (4.6)	**0.003**	196 (8.2)	188 (8.4)	8 (4.9)	**0.050**	187 (8.1)	163 (8.5)	24 (6.0)	0.228
Gender	Male	3370 (70.6)	3252 (70.7)	118 (67.8)		1728 (71.9)	1617 (72.3)	111 (67.6)		1653 (71.7)	1376 (72.2)	277 (69.8)	
	Female	1401 (29.4)	1345 (29.3)	56 (32.2)	0.406	674 (28.1)	621 (27.7)	53 (32.3)	0.209	651 (28.3)	531 (27.8)	120 (30.2)	0.338
Smoking status	Never	2447 (51.3)	2348 (51.1)	99 (56.9)		1264 (52.6)	1168 (52.2)	96 (58.5)		1222 (53.0)	993 (52.1)	229 (57.7)	
	Former (quit tobacco > 30days)	1190 (24.9)	1138 (24.8)	52 (29.9)		536 (22.3)	497 (22.2)	39 (23.8)		506 (22.0)	410 (21.5)	96 (24.2)	
	Current (tobacco < 30days)	1134 (23.8)	1111 (24.2)	23 (13.2)	**0.003**	602 (25.1)	573 (25.6)	29 (17.7)	0.076	576 (25.0)	504 (26.4)	72 (18.1)	**0.002**
History of dyslipidaemia		2349 (49.2)	2270 (49.4)	79 (45.4)	0.303	1142 (47.5)	1080 (48.3)	62 (37.8)	**0.010**	1091 (47.4)	919 (48.2)	172 (43.3)	0.077
History of hypertension		3565 (74.7)	3424 (74.5)	141 (81.0)	**0.051**	1800 (74.9)	1663 (74.3)	137 (83.5)	**0.008**	1720 (74.7)	1389 (72.8)	331 (83.4)	**<0.0001**
History of diabetes		2306 (48.3)	2218 (50.6)	88 (50.6)	0.547	1169 (48.7)	1079 (48.2)	90 (54.9)	0.099	1125 (48.8)	896 (47.0)	229 (57.7)	**<0.0001**
Family history of premature cardiovascular disease		630 (13.2)	614 (13.4)	16 (9.2)	0.111	280 (11.7)	273 (12.2)	7 (4.3)	**0.002**	270 (11.7)	247 (13.0)	23 (5.8)	**<0.0001**
History of MI/CAD		2244 (47.0)	2172 (47.2)	72 (41.4)	0.128	1131 (47.1)	1061 (47.4)	70 (42.7)	0.242	1079 (46.8)	890 (46.7)	189 (47.6)	0.734

Table 4.2, continued.

Variables	Description	In-hospital				30 days				1-year			
		Total	Survival	Non-survival	*p*-value	Total	Survival	Non-survival	*p*-value	Total	Survival	Non-survival	*p*-value
New onset angina (<2 weeks)		2646 (55.5)	2557 (55.6)	89 (51.1)	0.244	1214 (50.5)	1135 (50.7)	79 (6.5)	0.529	1161 (50.4)	992 (52.0)	169 (42.6)	**0.001**
History of heart failure		419 (8.8)	389 (8.5)	30 (17.2)	**<0.0001**	202 (8.4)	181 (8.1)	21 (12.8)	**0.036**	192 (8.3)	126 (6.6)	66 (16.6)	**<0.0001**
Chronic lung disease		243 (5.1)	226 (4.9)	17 (9.8)	**0.004**	140 (5.8)	128 (5.7)	12 (7.3)	0.399	133 (5.8)	101 (5.3)	32 (8.1)	**0.032**
Chronic renal disease		538 (11.3)	503 (10.9)	35 (20.1)	**<0.0001**	291 (12.1)	240 (10.7)	51 (31.1)	**<0.0001**	276 (12.0)	166 (8.7)	110 (27.7)	**<0.0001**
Cerebrovascular disease		227 (4.8)	211 (4.6)	16 (9.2)	0.005	121 (5.0)	107 (4.8)	14 (8.5)	0.034	118 (5.1)	82 (4.3)	36 (9.1)	**<0.0001**
Heart rate		84.04±20.41	83.54±20.04	97.09±25.28	**<0.0001**	85.35±20.96	84.61±20.50	95.46±24.36	**<0.0001**	85.21±20.78	83.85±20.25	91.72±22.06	**<0.0001**
Systolic blood pressure		144.05±27.66	144.48±27.7	132.69±34.59	**<0.0001**	143.44±27.65	144.13±27.16	134.03±32.38	**<0.0001**	143.87±27.43	144.22±26.59	142. ±31.15	0.175
Diastolic blood pressure		81.95±16.67	82.15±16.57	76.64±18.52	**<0.0001**	81.45±17.33	81.75±17.26	77.30±17.81	**0.001**	81.67±17.30	81.87±16.99	80.70±18.69	0.218
Killip class	I	3514 (73.7)	3455 (75.2)	59 (33.9)		1776 (73.9)	1710 (76.4)	66 (40.2)		1729 (75.0)	1526 (80.0)	203 (51.1)	
	II	851 (17.8)	808 (17.6)	43 (24.7)		373 (15.5)	328 (14.7)	45 (27.4)		341 (14.8)	235 (12.3)	106 (26.7)	
	III	279 (5.8)	249 (5.4)	30 (17.2)		171 (7.1)	149 (6.7)	22 (13.4)		162 (7.0)	110 (5.8)	52 (13.1)	
	IV	127 (2.7)	85 (1.8)	42 (24.1)	**<0.0001**	82 (3.4)	51 (2.3)	31 (18.9)	**<0.0001**	72 (3.1)	36 (1.9)	36 (9.1)	**<0.0001**
Total cholesterol		4.88±1.31	4.89±1.31	4.51±1.27	**<0.0001**	4.73±1.31	4.74±1.31	4.55±1.39	**0.073**	4.73±1.32	4.76±1.28	4.55±1.44	**0.003**
HDL		1.10±0.37	1.10±0.36	1.13±0.50	0.285	1.09±0.36	1.09±0.35	1.10±0.49	0.631	1.09±0.36	1.09±0.35	1.09±0.42	0.936
LDL		3.04±1.18	3.05±1.18	2.77±1.10	**0.002**	2.92±1.17	2.92±1.16	2.87±1.30	0.576	2.91±1.16	2.94±1.14	2.80±1.27	**0.037**
Fasting blood glucose		7.58±3.44	7.51±3.34	9.43±5.14	**<0.0001**	7.65±3.43	7.57±3.33	8.80±4.47	**<0.0001**	7.66±3.40	7.49±3.21	8.43±4.09	**<0.0001**
ECG abnormalities type	ST segment depression ≥0.5mm in ≥2 contiguous leads	1803 (37.8)	1692 (36.8)	111 (63.8)	**<0.0001**	908 (37.8)	820 (36.6)	88 (53.7)	**<0.0001**	856 (37.2)	665 (34.9)	191 (48.1)	**<0.0001**

Table 4.2, continued.

Variables	Description	In-hospital				30 days				1-year			
		Total	Survival	Non-survival	*p*-value	Total	Survival	Non-survival	*p*-value	Total	Survival	Non-survival	*p*-value
	T-wave inversion ≥1mm	2000 (41.9)	1945 (42.3)	55 (31.6)	**0.005**	1113 (46.3)	1050 (46.9)	63 (38.4)	**0.035**	1077 (46.7)	914 (47.9)	163 (41.1)	**0.013**
	Bundle branch block	295 (6.2)	278 (6.0)	17 (9.8)	**0.045**	136 (5.7)	117 (5.2)	19 (11.6)	**0.001**	128 (5.6)	86 (4.5)	42 (10.6)	**<0.0001**
ECG abnormalities location	Inferior leads: II, III, aVF	1084 (22.7)	1041 (22.6)	43 (24.7)	0.523	552 (23.0)	520 (23.2)	32 (19.5)	0.274	529 (23.0)	454 (23.8)	75 (18.9)	**0.034**
	Anterior leads: V1 to V4	1734 (36.3)	1655 (36.0)	79 (45.4)	**0.011**	872 (36.3)	799 (35.7)	73 (44.5)	**0.024**	835 (36.2)	666 (34.9)	169 (42.6)	**0.004**
	Lateral leads: I, aVL, V5 to V6	1921 (40.3)	1840 (40.0)	81 (46.6)	0.085	1061 (44.2)	975 (43.6)	86 (52.4)	**0.027**	1019 (44.2)	809 (42.4)	210 (52.9)	**<0.0001**
Cardiac catheterization		1310 (27.5)	1280 (27.8)	30 (17.2)	**0.002**	665 (27.7)	630 (28.2)	35 (21.3)	0.060	621 (27.0)	557 (29.2)	64 (16.1)	**<0.0001**
PCI		719 (15.1)	703 (15.3)	16 (9.2)	**0.027**	360 (15.0)	341 (15.2)	19 (11.6)	0.206	344 (14.9)	316 (16.6)	28 (7.1)	**<0.0001**
LMWH		2291 (48.0)	2190 (47.6)	101 (58.0)	**0.007**	651 (27.1)	563 (25.2)	88 (53.7)	**<0.0001**	596 (25.9)	425 (22.3)	171 (43.1)	**<0.0001**
Beta blockers		3093 (64.8)	3025 (65.8)	68 (39.1)	**<0.0001**	1398 (58.2)	1321 (59.0)	77 (47.0)	0.002	1321 (57.3)	1090 (57.20	231 (58.2)	**<0.0001**
ACE inhibitors		2530 (53.0)	2487 (54.1)	43 (24.7)	**<0.0001**	1060 (44.1)	1018 (45.5)	42 (25.6)	**<0.0001**	1008 (43.8)	877 (46.0)	131 (33.0)	**<0.0001**
Angiotensin II receptor blocker		436 (9.1)	431 (9.4)	5 (2.9)	**0.003**	195 (8.1)	188 (8.4)	7 (4.3)	0.061	186 (8.1)	166 (8.7)	20 (5.0)	**0.015**
Statin		4464 (93.6)	4316 (93.9)	148 (85.1)	**<0.0001**	2255 (93.9)	2110 (94.3)	145 (88.4)	**0.002**	2162 (93.8)	1795 (94.1)	367 (92.4)	0.204
Diuretics		1546 (32.4)	1446 (31.5)	100 (57.5)	**<0.0001**	838 (34.9)	741 (33.1)	97 (59.1)	**<0.0001**	783 (34.0)	564 (29.6)	219 (55.2)	**<0.0001**
Calcium antagonist		1111 (23.3)	1081 (23.5)	30 (17.2)	0.055	592 (24.6)	557 (24.9)	35 (21.3)	0.309	578 (25.1)	477 (25.0)	101 (25.4)	0.858
Oral hypoglycaemic agent		1429 (30.0)	1412 (30.7)	17 (9.8)	**<0.0001**	692 (28.8)	669 (29.9)	23 (14.0)	**<0.0001**	672 (29.2)	577 (30.3)	95 (23.9)	**0.012**
Insulin		1081 (22.7)	1029 (22.4)	52 (29.9)	**0.020**	591 (24.6)	535 (23.9)	56 (34.1)	**0.003**	566 (24.6)	435 (22.8)	131 (33.0)	**<0.0001**

p-value is statistically highly significant as p<0.001

4.2 ML prediction

Maximal predictive performances on the 30% untouched testing dataset were observed for ML models constructed using complete and reduced sets of variables compared to the TIMI risk score. The criteria for finding the best model among all the ML algorithms were the model with the highest calibration value (AUC) and the least number of variables.

As for the STEMI dataset, all of the ML models outperformed the TIMI risk score across the three time points (Table 4.3). The best-selected ML model (SVMvarImp–SBE–SVM) also performed better against TIMI based on the AUC value using the untouched 30% testing dataset ($p < 0.0001$ for all models). Detailed performance evaluation of the best ML model against TIMI risk score is presented in Table 4.4.

Table 4.3: The AUC of TIMI risk score and ML models with and without feature selection based on 30% testing STEMI dataset.

Classifiers	The area under the ROC Curve (95% CI)		
	In-hospital	30 days	1-year
RF	0.86 (0.820-0.88)	0.83 (0.786-0.879)	0.78 (0.741-0.827)
RFvarImp-SBE-RF	0.87 (0.832-0.907)	0.85 (0.10-0.890)	0.80 (0.750-0.834)
RFE-RF	0.86 (0.821-0.893)	0.82 (0.772-0.872)	0.79 (0.748-0.833)
SVM	0.86 (0.824-0.895)	0.87 (0.831-0.912)	0.84 (0.801-0.877)
SVMvarImp-SBE-SVM	0.88 (0.846-0.910)	0.90 (0.870-0.935)	0.84 (0.798-0.872)

Table 4.3, continued.

Classifiers	The area under the ROC Curve (95% CI)		
	In-hospital	30 days	1-year
RFE-SVM	0.85 (0.811-0.887)	0.88 (0.837-0.920)	0.84 (0.806-0.880)
LR	0.88 (0.846-0.911)	0.85 (0.803-0.897)	0.76 (0.710-0.807)
LRstepwise - SBE-LR	0.89 (0.861-0.920)	0.85 (0.812-0.906)	0.80 (0.767-0.848)
RFE- LR	0.87 (0.842-0.897)	0.83 (0.783-0.882)	0.78 (0.737-0.826)
TIMI	0.81 (0.772-0.802)	0.80 (0.746-0.838)	0.76 (0.715-0.802)

Table 4.4: Additional performance metrics based on 30% STEMI testing dataset for TIMI risk score and ML models with and without feature selection.

	PPV	NPV	Sensitivity	Specificity	Accuracy (Cl 95%)	Mcnemar's test (p-value)
In-hospital						
Classifier						
RF	0.380	0.963	0.347	0.968	0.935 (0.923,0.946)	<0.0001
RFvarImp-SBE-RF	0.447	0.963	0.337	0.977	0.942 (0.931,0.952)	<0.0001
RFE-RF	0.350	0.963	0.347	0.964	0.931 (0.918,0.942)	<0.0001
SVM	0.242	0.976	0.614	0.892	0.877 (0.861, 0.891)	<0.0001
SVMvarImp-SBE-SVM	0.219	0.980	0.693	0.861	0.852 (0.835,0.868)	<0.0001
RFE-SVM	0.202	0.982	0.723	0.838	0.832 (0.815, 0.849)	<0.0001
LR	0.211	0.981	0.713	0.850	0.842 (0.825,0.858)	<0.0001
LRstepwise - SBE-LR	0.211	0.984	0.752	0.841	0.836 (0.814,0.852)	<0.0001

Table 4.4, continued.

	PPV	NPV	Sensitivity	Specificity	Accuracy (Cl 95%)	Mcnemar's test (p-value)
In-hospital						
Classifier						
RFE- LR	0.185	0.978	0.663	0.834	0.825 (0.807,0.842)	<0.0001
TIMI	0.180	0.976	0.644	0.834	0.824 (0.806, 0.841)	<0.0001
30 days						
Classifier						
RF	0.389	0.946	0.373	0.949	0.903 (0.882,0.921)	<0.0001
RFvarImp-SBE-RF	0.341	0.948	0.413	0.930	0.889 (0.867,0.909)	<0.0001
RFE-RF	0.414	0.947	0.387	0.952	0.907 (0.887,0.925)	<0.0001
SVM	0.258	0.972	0.733	0.817	0.810 (0.784,0.835)	<0.0001
SVMvarImp-SBE-SVM	0.258	0.983	0.840	0.790	0.794 (0.767,0.820)	<0.0001

Table 4.4, continued.

	PPV	NPV	Sensitivity	Specificity	Accuracy (Cl 95%)	Mcnemar's test (p-value)
30 days						
Classifiers						
RFE-SVM	0.261	0.980	0.813	0.800	0.801 (0.774,0.829)	<0.0001
LR	0.248	0.973	0.747	0.803	0.799 (0.771,0.824)	<0.0001
LRstepwise - SBE-LR	0.281	0.974	0.747	0.834	0.827 (0.802,0.851)	<0.0001
RFE- LR	0.248	0.971	0.720	0.810	0.803 (0.776,0.828)	<0.0001
TIMI	0.245	0.962	0.627	0.832	0.816 (0.789, 0.840)	<0.0001
1-year						
Classifier						
RF	0.410	0.909	0.373	0.949	0.827 (0.801,0.852)	<0.0001
RFvarImp-SBE-RF	0.436	0.909	0.460	0.901	0.838 (0.811,0.861)	<0.0001

Table 4.4, continued.

	PPV	NPV	Sensitivity	Specificity	Accuracy (Cl 95%)	Mcnemar's test (p-value)
1-year						
Classifier						
RFE-RF	0.425	0.913	0.492	0.889	0.8318 (0.805,0.856)	<0.0001
SVM	0.382	0.950	0.746	0.798	0.7909 (0.763,0.817)	<0.0001
SVMvarImp-SBE-SVM	0.357	0.950	0.754	0.773	0.771 (0.741,0.798)	<0.0001
RFE-SVM	0.387	0.953	0.762	0.798	0.793 (0.765,0.820)	<0.0001
LR	0.329	0.924	0.611	0.792	0.766 (0.737,0.794)	<0.0001
LRstepwise - SBE-LR	0.372	0.935	0.659	0.814	0.792 (0.764,0.818)	<0.0001
RFE- LR	0.344	0.926	0.619	0.802	0.776 (0.747,0.803)	<0.0001

Table 4.4, continued.

	PPV	NPV	Sensitivity	Specificity	Accuracy (Cl 95%)	Mcnemar's test (p-value)
1-year						
Classifier						
TIMI	0.332	0.907	0.484	0.837	0.786 (0.758, 0.813)	<0.0001

Based on the criteria of the best selected model to be employed as mentioned above, the combination of SVMvarImp-SBE-SVM algorithm demonstrated the highest predictive performance with the least number of predictors for in-hospital, 30 days and 1-year models. Figure 4.1(a) illustrates all ML models' performances and Figure 4.1(b) illustrates the best selected ML models against TIMI risk score performances based on the AUC value using the untouched 30% validation dataset. There was no significant difference in the in-hospital model for LRstepwise-SBE-LR (AUC = 0.89, 95% CI:0.861–0.920) with 24 variables and SVMvarImp-SBE-SVM (AUC = 0.88, 95% CI: 0.846–0.910) with 15 variables (p = 0.143; 95% CI, -0.026 to 0.004). 30-days ML mortality prediction for SVMvarImp–SBE–SVM (AUC = 0.90, 95% CI: 0.867–0.935) model also reported no significant difference to RFE-SVM (AUC = 0.88,95% CI:0.837–0.920) (p = 0.115; 95% CI, -0.013 to 0.001). Model performances were observed to be similar (AUC = 0.84) and showed a significant difference (p <0.0001) between the following 1-year mortality models (SVMvarImp–SBE–SVM vs SVM; 95% CI, 0.035 to 0.052, RFE-SVM vs SVM, 95% CI, 0.005 to 0.011, SVMvarImp-SBE-SVM vs RFE-SVM; 95% CI, 0.027 to 0.044). However, SVMvarImp-SBE-SVM model consisted of the least number of variables (12 variables) compared to SVM without feature selection (50 variables) and RFE-SVM (44 variables). Similar performance was also reported for LR with a reduced set of predictors (AUC = 0.85, 95% CI: 0.812–0.907) and a complete set of predictors for 30 days (AUC = 0.85, 95% CI: 0.803–0.897) but showed no significant difference (p = 0.828; 95% CI, -0.007 to 0.009).

For the secondary analysis, missing values in each variable were imputed and utilised as the training dataset for model development, with 30% untouched testing dataset used for model performance evaluation. Secondary analysis on best ML models (SVMvarImp–SBE–SVM) performance trained with imputed data reported for in-hospital (AUC = 0.87,

95% CI: 0.845–0.912), 30 days (AUC = 0.90, 95% CI: 0.857–0.923), and 1-year (AUC = 0.83, 95% CI: 0.796–0.871). Complete case ML model dataset resulted in an almost similar AUC result for in-hospital (AUC = 0.88, 95% CI: 0.846–0.910), 30 days (AUC = 0.90, 95% CI: 0.870–0.935), and 1-year mortality (AUC = 0.84, 95% CI: 0.798–0.872). For in-hospital and 30 days, the imputed and the complete case model was significant ($p<0.0001$; 95% CI: 0.011 to 0.018, $p = 0.001$; 95% CI: 0.004 to 0.016 respectively). As for 1-year model, it was not statistically significant between imputed and complete case model ($p = 0.931$; 95% CI: -0.006 to 0.005).

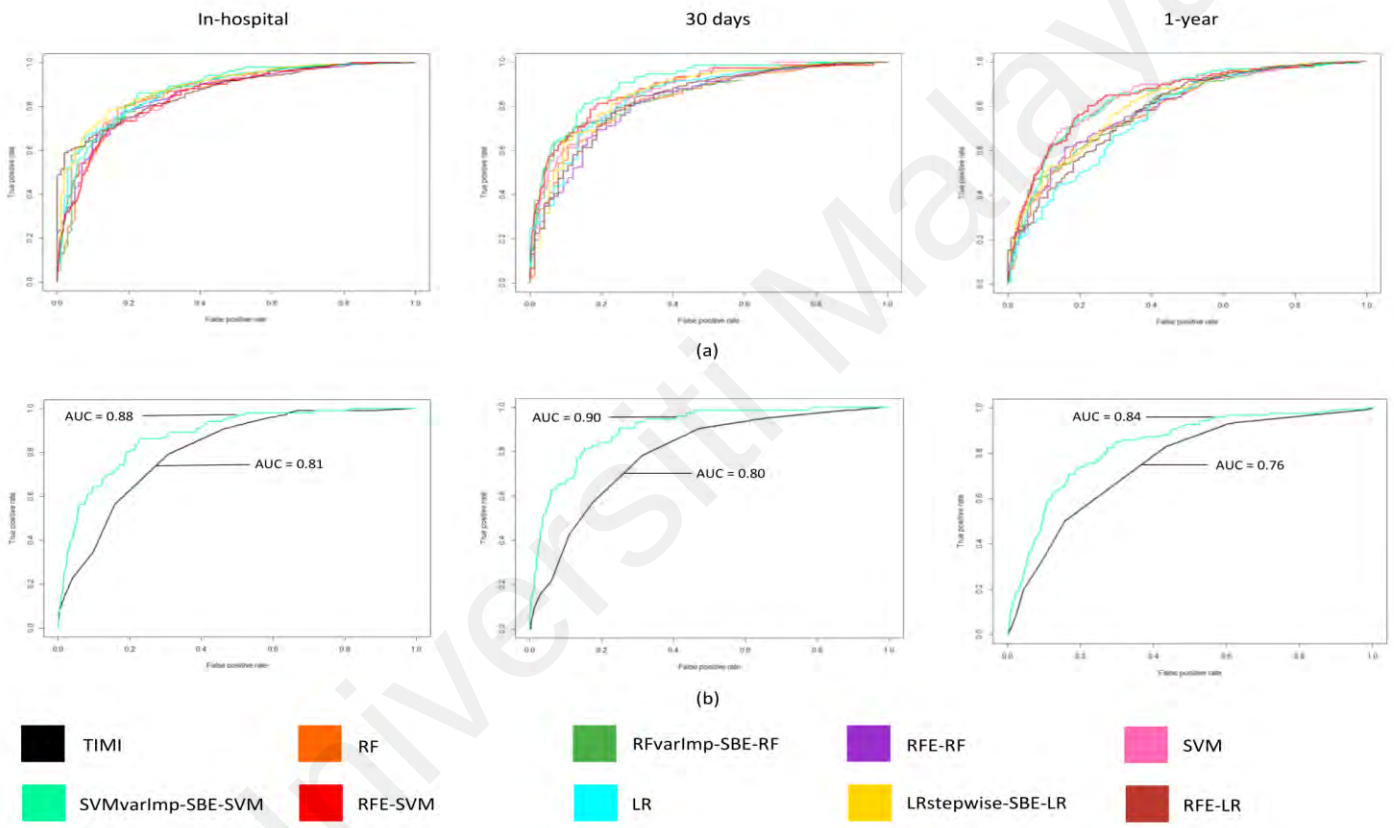


Figure 4.1:The receiver operating characteristics (ROC) curves of ML models and TIMI score are based on a 30% testing STEMI dataset. ROC curves show the performance for In-hospital, 30 days and 1-year ML mortality prediction models (a). The ROC values for TIMI against the best ML model (SVMvarImp-SBS-SVM).

On the other hand, the result for the NSTEMI/UA dataset showed that all ML models (reduced and complete dataset) performed better than the TIMI risk score for all three different time points of mortality prediction using the 30% of the untouched testing dataset (Table 4.5). Based on the AUC value and the number of selected variables, SVMvarImp–SBE–SVM was selected as the best model for all three time points for the NSTEMI/UA dataset and it performed better against the TIMI risk score based on the AUC value utilising the untouched 30% testing dataset, similar to the STEMI dataset ($p < 0.0001$ for all models). Table 4.6 shows the detailed performance evaluation of the best ML models against the TIMI risk score.

Table 4.5: The AUC of TIMI risk score and ML models with and without feature selection based on 30% testing NSTEMI/UA dataset.

Classifiers	The area under the ROC Curve (95% CI)		
	In-hospital	30 days	1-year
RF	0.81 (0.762-0.867)	0.81 (0.757-0.859)	0.80 (0.753-0.838)
RFvarImp-SBE-RF	0.82 (0.764-0.870)	0.85 (0.812-0.890)	0.80 (0.754-0.838)
RFE-RF	0.79 (0.737-0.848)	0.78 (0.728-0.841)	0.79 (0.748-0.835)
SVM	0.80 (0.745-0.860)	0.82 (0.765-0.873)	0.80 (0.756-0.837)
SVMvarImp-SBE-SVM	0.85 (0.81-0.889)	0.87 (0.824-0.907)	0.80 (0.756-0.840)
RFE-SVM	0.76 (0.693-0.827)	0.82 (0.775-0.870)	0.79 (0.749-0.832)
LR	0.81 (0.755-0.869)	0.80 (0.733-0.860)	0.77 (0.724-0.815)

Table 4.5, continued.

Classifiers	The area under the ROC Curve (95% CI)		
	In-hospital	30 days	1-year
LRstepwise - SBE-LR	0.82 (0.764-0.871)	0.83 (0.783-0.883)	0.78 (0.736-0.824)
RFE- LR	0.76 (0.689-0.827)	0.84 (0.788-0.886)	0.78 (0.733-0.820)
TIMI	0.42 (0.356-0.494)	0.49 (0.410-0.561)	0.42 (0.368-0.477)

Table 4.6: Additional performance metrics based on 30% NSTEMI/UA testing dataset for TIMI risk score and ML models with and without feature selection.

	PPV	NPV	Sensitivity	Specificity	Accuracy (Cl 95%)	Mcnemar's test (p-value)
In-hospital						
Classifier						
RF	0.455	0.967	0.096	0.996	0.963 (0.9518,0.9721)	<0.0001
RFvarImp-SBE-RF	0.238	0.970	0.192	0.977	0.948 (0.936,0.959)	<0.0001
RFE-RF	0.310	0.969	0.173	0.985	0.956 (0.944,0.966)	<0.0001
SVM	0.120	0.979	0.519	0.856	0.844 (0.824,0.863)	<0.0001
SVMvarImp-SBE-SVM	0.122	0.985	0.673	0.817	0.812 (0.791,0.832)	<0.0001
RFE-SVM	0.094	0.978	0.519	0.811	0.800 (0.779,0.821)	<0.0001
LR	0.117	0.983	0.615	0.825	0.818 (0.797,0.837)	<0.0001

Table 4.6, continued.

	PPV	NPV	Sensitivity	Specificity	Accuracy (Cl 95%)	Mcnemar's test (p-value)
30 days						
Classifier						
LRstepwise - SBE-LR	0.108	0.981	0.577	0.819	0.811 (0.789,0.831)	<0.0001
RFE- LR	0.107	0.982	0.596	0.812	0.804 (0.783,0.825)	<0.0001
TIMI	0.031	0.963	0.077	0.910	0.879 (0.862,0.896)	<0.0001
30 days						
Classifier						
RF	0.278	0.937	0.102	0.981	0.921 (0.899,0.940)	<0.0001
RFvarImp-SBE-RF	0.318	0.940	0.143	0.978	0.921 (0.899,0.940)	<0.0001
RFE-RF	0.217	0.945	0.265	0.930	0.885 (0.859,0.907)	<0.0001
SVM	0.193	0.971	0.673	0.794	0.786 (0.754,0.816)	<0.0001

Table 4.6, continued.

	PPV	NPV	Sensitivity	Specificity	Accuracy (Cl 95%)	Mcnemar's test (p-value)
30 days						
Classifier						
SVMvarImp-SBE-SVM	0.246	0.974	0.694	0.845	0.835 (0.806,0.861)	<0.0001
RFE-SVM	0.157	0.967	0.653	0.744	0.738 (0.704,0.769)	<0.0001
LR	0.173	0.971	0.694	0.759	0.754 (0.721,0.785)	<0.0001
LRstepwise - SBE-LR	0.200	0.976	0.735	0.785	0.782 (0.750,0.812)	<0.0001
RFE- LR	0.183	0.974	0.714	0.768	0.764 (0.731,0.795)	<0.0001
TIMI	0.058	0.931	0.102	0.879	0.826 (0.797,0.853)	<0.0001
1-year						
Classifier						

Table 4.6, continued.

	PPV	NPV	Sensitivity	Specificity	Accuracy (Cl 95%)	Mcnemar's test (p-value)
1-year						
Classifier						
RF	0.500	0.881	0.403	0.916	0.828 (0.798,0.855)	<0.0001
RFvarImp-SBE-RF	0.429	0.893	0.504	0.860	0.799 (0.767,0.828)	<0.0001
RFE-RF	0.491	0.887	0.445	0.904	0.825 (0.795,0.853)	<0.0001
SVM	0.369	0.905	0.605	0.785	0.754 (0.720,0.786)	<0.0001
SVMvarImp-SBE-SVM	0.333	0.945	0.815	0.661	0.687 (0.651,0.722)	<0.0001
RFE-SVM	0.372	0.909	0.622	0.782	0.754 (0.720,0.786)	<0.0001
LR	0.354	0.915	0.664	0.748	0.734 (0.699,0.766)	<0.0001
LRstepwise - SBE-LR	0.358	0.913	0.656	0.755	0.738 (0.704,0.771)	<0.0001

Table 4.6, continued.

	PPV	NPV	Sensitivity	Specificity	Accuracy (Cl 95%)	Mcnemar's test (p-value)
1-year						
Classifier						
RFE- LR	0.346	0.906	0.622	0.755	0.732 (0.698,0.765)	<0.0001
TIMI	0.224	0.834	0.143	0.897	0.767 (0.734,0.798)	<0.0001

Figure 4.2(a) shows the performance of all ML models in the NSTEMI/UA dataset, whereas Figure 4.2(b) shows the performance of the best-selected ML models against the TIMI risk score for all three time points based on their AUC values using the 30% untouched testing dataset. As seen from the graphs, the combination of SVMvarImp–SBE–SVM for in-hospital, 30 days, and 1 year shows the best performance among other ML models and TIMI risk score. In the 30 days mortality prediction, SVMvarImp–SBE–SVM performed significantly better ($p<0.0001$, 95% CI: 0.144 to 0.169) than RFvarImp-SBE-RF even though there was not much difference in the AUC values (AUC=0.87, 95% CI: 0.824-0.907 with 11 variables and AUC=0.85, 95% CI: 0.812-0.890 with 16 variables respectively). In the 1-year mortality prediction, the SVMvarImp–SBE–SVM (AUC=0.80, 95% CI: 0.756-0.840) model reported significantly better performance as compared to the RFE-SVM (AUC=0.79, 95% CI:0.749-0.832) and RFE-RF (AUC=0.79, 95% CI: 0.748-0.835) with a slightly higher value of AUC ($p<0.0001$, SVMvarImp–SBE–SVM vs RFE-SVM; 95% CI: 0.014 to 0.031, SVMvarImp–SBE–SVM vs RFE-RF; 95% CI: 0.072 to 0.0.89). On the other hand, a few models also showed similar performances (AUC=0.80) and highly significant ($p<0.0001$) between the 1-year mortality prediction models between SVMvarImp–SBE–SVM vs RF (95% CI,0.069 to 0.086), SVMvarImp–SBE–SVM vs RFvarImp–SBE–RF (95% CI: 0.073 to 0.091), SVMvarImp–SBE–SVM vs SVM (95% CI: 0.023 to 0.041), SVM vs RF (95% CI: 0.035 to 0.055), and SVM vs RFvarImp–SBE–RF (95% CI: 0.039 to 0.061). RFvarImp–SBE–RF vs RF were the only models' comparison that had the same AUC values but has no significance ($p=0.151$,95% CI: -0.011 to 0.002). However, The SVMvarImp-SBE-SVM model has the smallest number of predictors (13 variables) as compared to the RF (39 variables), SVM (39 variables), and RFvarImp–SBE–RF (17 variables).

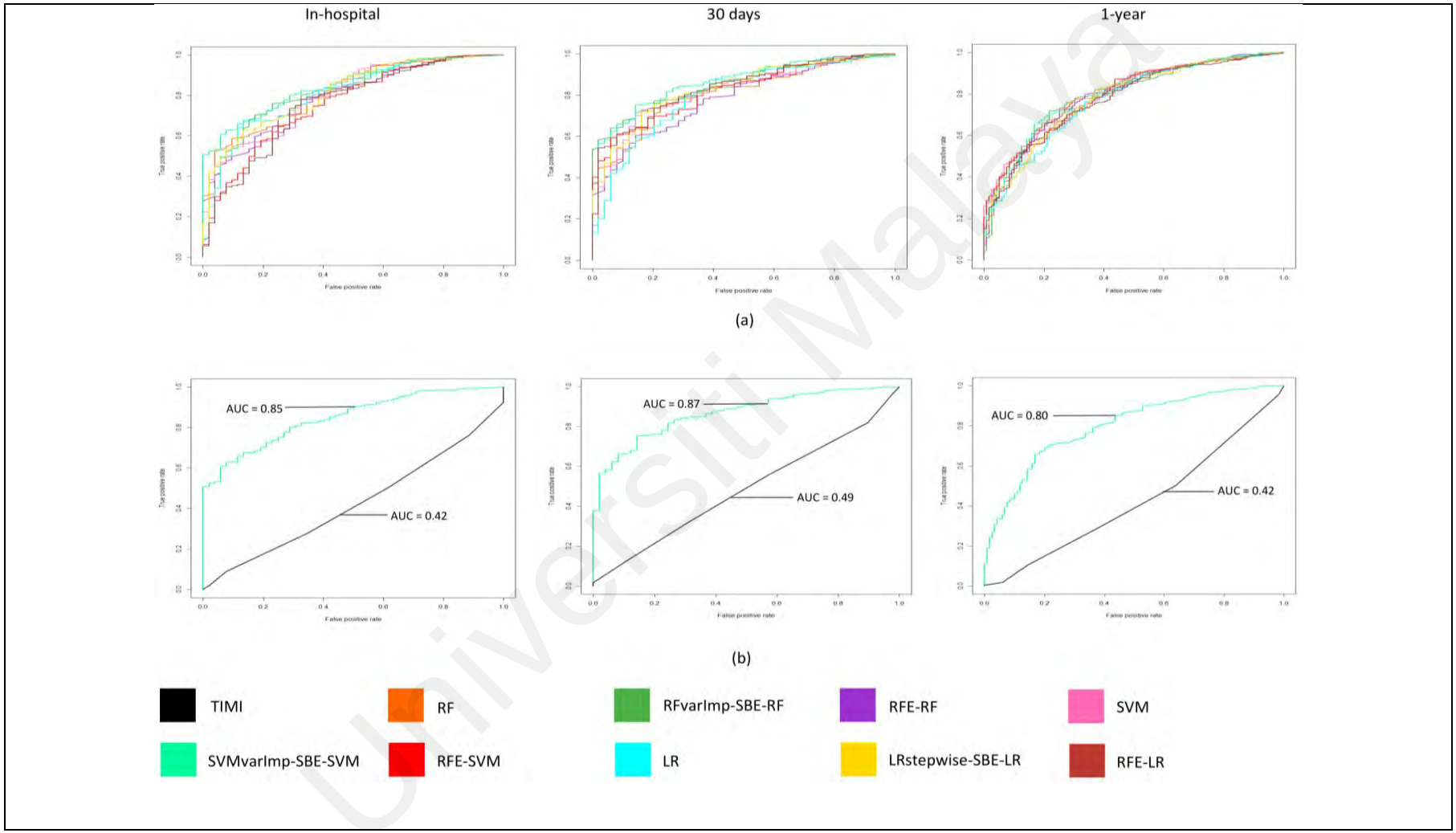


Figure 4.2: **The receiver operating characteristics (ROC) curves of ML models and TIMI score are based on a 30% testing NSTEMI/UA dataset. (a) ROC curves show the performance for In-hospital, 30 days and 1-year ML mortality prediction models. The ROC values for TIMI against the best ML model (SVMvarImp-SBS-SVM).**

For the three time points of mortality prediction, a secondary analysis was conducted on the best ML model (SVMvarImp–SBE–SVM) and the imputed data were used as the training dataset. The performance of the imputed dataset for the NSTEMI/UA reported with an AUC of 0.81 (95% CI: 0.762 to 0.862) for in-hospital, 0.86 (95% CI: 0.817 to 0.907) for 30 days and 0.80 (95% CI: 0.754 to 0.833) for 1-year mortality prediction. The AUC values of the imputed dataset and the complete dataset were almost similar and highly significant for in-hospital, 30 days and 1-year ($p<0.0001$; 95% CI: 0.010 to 0.028, 95% CI: 0.016 to 0.027, 95% CI: 0.019 to 0.028 respectively).

4.3 Feature selection

RFE and SBE feature selection methods were combined with ML algorithms to construct predictive models with optimal performance (refer to Methodology). Initial ranking using all 50 (for STEMI) and 39 (for NSTEMI/UA) variables for best model (SVMvarimp-SBE-SVM) using SVM variable importance was shown in Figures 3.8 and 3.10 in Chapter 3. SBE was then used to identify features that result in model optimal performances.

For the STEMI dataset, the common predictors observed for in-hospital, 30 days, and 1-year mortality for all ML models in this study were age, heart rate, Killip class, and fasting blood glucose (Table 4.7). Diuretics intake was an additional common predictor for the best model (SVMvarImp-SBE-SVM). Age, heart rate, and Killip class were identified as common predictors for the best ML model (SVMvarImp- SBE-SVM) for in-hospital, 30 days, and 1-year against TIMI (Table 4.8).

Table 4.7: The selected variables using RFE and SBE for all the ML models in the STEMI dataset for in-hospital, 30 days and 1-year mortality prediction.

Variables	In-hospital				30 days				1 year			
	RF-SBE	SVM-SBE	RFE	SBE- LR	RF-SBE	SVM-SBE	RFE	SBE-LR	RF-SBE	SVM-SBE	RFE	SBE-LR
Age	•	•	•	•	•	•	•	•	•	•	•	•
Race	•			•		•		•			•	•
Sex											•	•
Smoking status	•									•	•	
Hypertension				•	•					•	•	
Diabetes										•	•	
Family history of premature CVD		•		•		•					•	
Documented CAD				•							•	•
Heart failure								•			•	
Chronic renal disease		•						•	•		•	•
Chronic lung disease											•	
Heart rate	•	•	•	•	•	•	•	•	•	•	•	•
Systolic bp	•	•	•	•			•	•	•	•	•	•
Diastolic bp	•	•	•		•						•	
Killip class	•	•	•	•	•	•	•	•	•	•	•	•
Total cholesterol	•		•		•		•	•			•	

Table 4.7, continued.

Variables	In-hospital				30 days				1 year			
	RF-SBE	SVM-SBE	RFE	SBE- LR	RF-SBE	SVM-SBE	RFE	SBE-LR	RF-SBE	SVM-SBE	RFE	SBE-LR
HDL	•			•		•		•	•		•	•
LDL	•		•	•					•		•	•
Triglycerides	•				•						•	
Fasting blood glucose	•	•	•	•		•	•	•	•	•	•	•
ECG-type elevation ≥ 1mm											•	
ECG-type elevation ≥ 2mm											•	
ECG-type depression ≥ 0.5mm											•	
ECG- type T-wave								•				
ECG-type bundle branch block		•		•				•			•	•
ECG- location inferior lead				•							•	
ECG- location anterior lead					•			•			•	•
ECG- location lateral lead	•	•		•				•	•		•	•
FB status				•							•	

Table 4.7, continued.

Variables	In-hospital				30 days				1 year			
	RF-SBE	SVM-SBE	RFE	SBE- LR	RF-SBE	SVM-SBE	RFE	SBE-LR	RF-SBE	SVM-SBE	RFE	SBE-LR
Cardiac catheterization		•		•	•	•		•			•	•
PCI						•				•	•	
CABG				•								
ASA				•		•		•			•	
GP receptor inhibitor											•	
Heparin				•							•	
LMWH								•			•	•
Beta-blockers	•	•		•	•	•	•	•	•		•	•
ACE inhibitors	•			•						•	•	•
Angiotensin II receptor blocker											•	
Statin		•									•	•
Other lipid-lowering agents											•	
Diuretics		•		•	•	•		•	•	•	•	•
Calcium antagonists								•			•	
Oral hypoglycaemic agent	•	•	•	•		•	•	•			•	

Table 4.7, continued.

Variables	In-hospital				30 days				1 year			
	RF-SBE	SVM-SBE	RFE	SBE- LR	RF-SBE	SVM-SBE	RFE	SBE-LR	RF-SBE	SVM-SBE	RFE	SBE-LR
Insulin					•					•	•	
Anti-arrhythmic agent	•			•				•			•	

Table 4.8: Selected variables that resulted in optimum AUC for the best ML models (SVMvarImp-SBE-SVM) in in-hospital, 30-days, and 1-year against TIMI risk score for STEMI variables.

Variables	Machine Learning best model			TIMI Score
	In-hospital	30 days	1-year	
Age	•	•	•	•
Race		•		
Smoking status			•	
Hypertension			•	•
Diabetes			•	•
Family history of premature CVD	•	•		
Chronic renal disease	•			
Heart rate	•	•	•	•

Table 4.8, continued.

Variables	Machine Learning best model			TIMI score
	In-hospital	30 days	1-year	
Systolic bp	•		•	•
Diastolic bp	•			
Killip class	•	•	•	•
HDL		•		
Fasting blood glucose	•	•	•	
Weight				•
ECG-type bundle branch block	•			•
ECG- location lateral lead	•			
Time to treatment				•
Cardiac catheterization	•	•		
PCI		•	•	
ASA		•		
Beta-blockers	•	•		
ACE inhibitors			•	
Statin	•			

Table 4.8, continued.

Variables	Machine Learning best model			TIMI score
	In-hospital	30 days	1-year	
Diuretics	•	•	•	
Oral hypoglycaemic agent	•	•		
Insulin			•	

The selected variables for the NSTEMI/UA dataset using RFE and SBE can be referred to in Table 4.9 below. The common variables across all ML models were age and Killip class for in-hospital, 30 days, and 1-year mortality prediction. Including those two, heart rate and Low-Molecular-Weight Heparin (LMWH) intake were the common predictors of the best ML models (SVMvarImp-SBE-SVM) for the three-time points. Table 4.10 below shows the variables of the best ML models for in-hospital, 30 days and 1 year against TIMI score for NSTEMI/UA and age was the only common variable between them.

Table 4.9:The selected variables using RFE and SBE for all the ML models in the NSTEMI/UA dataset for in-hospital, 30 days and 1-year mortality prediction.

Variables	In-hospital				30 days				1 year			
	RF-SBE	SVM-SBE	SBE - LR	RFE	RF-SBE	SVM-SBE	SBE - LR	RFE	RF-SBE	SVM-SBE	SBE - LR	RFE
Age	•	•	•	•	•	•	•	•	•	•	•	•
Sex	•	•	•								•	•
Race												
Smoking status	•		•						•			•
Dyslipidaemia	•						•				•	•
Diabetes				•								•
Hypertension	•	•	•				•				•	
Family history of premature CVD	•								•		•	
History of MI/CAD	•	•			•							•
Chronic angina (≥2 weeks)			•						•	•		
Heart failure				•	•			•	•		•	•
Chronic lung disease		•							•			•
Chronic renal disease					•	•	•	•	•	•	•	•
Cerebrovascular disease									•		•	•
Heart rate	•	•	•		•	•	•			•	•	•
Systolic bp	•	•	•	•	•	•	•	•	•		•	•
Diastolic bp				•	•				•			•

Table 4.9, continued.

Variables	In-hospital				30 days				1 year			
	RF-SBE	SVM-SBE	SBE - LR	RFE	RF-SBE	SVM-SBE	SBE - LR	RFE	RF-SBE	SVM-SBE	SBE - LR	RFE
Killip class	•	•	•	•	•	•	•	•	•	•	•	•
Total cholesterol				•	•		•	•				•
HDL					•	•		•	•			•
LDL				•	•		•	•		•		•
Fasting blood glucose	•	•	•	•	•		•		•		•	•
ECG-type depression ≥ 0.5mm		•	•						•	•		•
ECG- type T-wave							•				•	
ECG-type bundle branch block						•		•		•	•	•
ECG- location inferior lead						•						•
ECG- location anterior lead	•										•	•
ECG- location lateral lead					•					•	•	•
PCI			•							•	•	•
Cardiac catheterization						•			•	•		•
LMWH	•	•	•		•	•	•			•	•	•
Beta-blockers	•	•	•	•	•		•				•	•
ACE inhibitors			•								•	•

Table 4.9, continued.

Variables	In-hospital				30 days				1 year			
	RF-SBE	SVM-SBE	SBE - LR	RFE	RF-SBE	SVM-SBE	SBE - LR	RFE	RF-SBE	SVM-SBE	SBE - LR	RFE
Angiotensin II receptor blocker		•	•				•				•	•
Statin			•				•					
Diuretics			•					•	•	•	•	•
Calcium antagonists	•			•							•	•
Oral hypoglycaemic agent	•		•	•	•	•	•	•	•		•	•
Insulin			•									•

Table 4.10: Selected variables that resulted in optimum AUC for the best ML models (SVMvarImp-SBE-SVM) in in-hospital, 30 days and 1-year against TIMI risk score for NSTEMI/UA variables.

Variables	Machine learning best model			TIMI score
	In-hospital	30 days	1-year	
Age	•	•	•	•
Gender	•			
Smoking status				•
Diabetes mellitus				
Hypertension	•			
Family history of premature CVD				
History	•			
Known CAD (stenosis ≥50%)				•
Severe angina (≥2 episodes in 24 hrs)				•
Chronic angina (≥ 2 weeks)			•	
Chronic lung disease	•			
Chronic renal disease		•	•	

Table 4.10, continued.

Variables	Machine learning best model			TIMI score
	In-hospital	30 days	1-year	
Heart rate	•	•	•	
Systolic bp	•	•		
Killip class	•	•	•	
HDL		•		
LDL			•	
Fasting blood glucose	•			
Positive cardiac marker				•
ECG-type depression ≥ 0.5 mm	•		•	•
ECG-type bundle branch block		•	•	
ECG-location inferior lead		•		
ECG-location lateral lead			•	
PCI			•	
Cardiac catheterization		•	•	
Aspirin				•

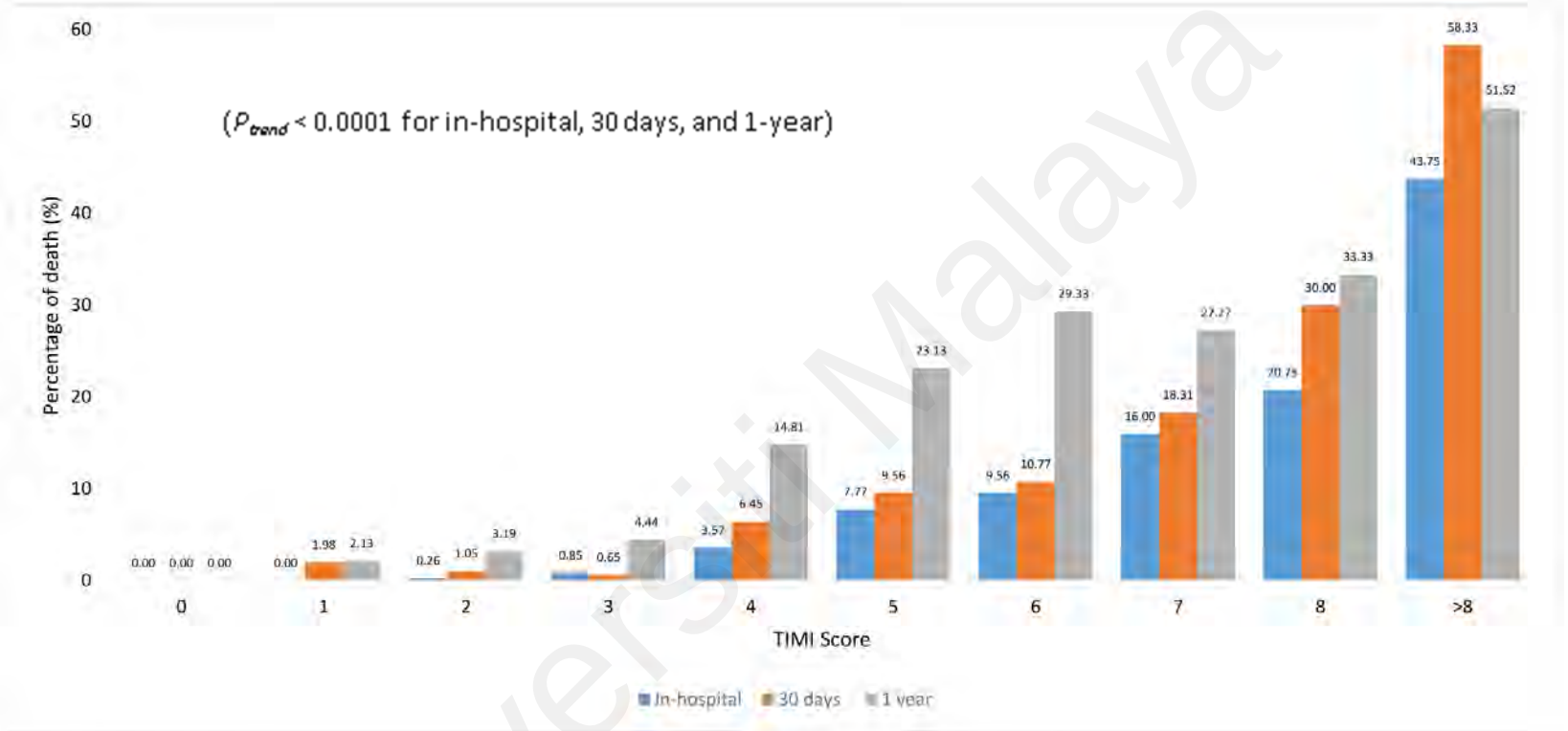
Table 4.10, continued.

Variables	Machine learning best model			TIMI score
	In-hospital	30 days	1-year	
LMWH	•	•	•	
Beta-blockers	•			
Angiotensin II receptor blocker	•			
Diuretics			•	
Oral hypoglycaemic agent		•		

4.4 Comparison of ML to TIMI risk score when applied to the validation dataset

After the best ML models were identified for the three time points, the mortality rate of each probability percentile (from 0% to 100%) of the ML models and mortality prediction were calculated. Figures 4.3 and 4.4 illustrate the comparison of the best ML model mortality rate against the TIMI risk score for the STEMI dataset and Figures 4.5 and 4.6 for the NSTEMI/UA dataset. As mentioned in Chapter 2, TIMI score for STEMI categorises patients as low risk at the score of ≤5 and a high-risk score of > 5 (Morrow et al., 2000) while TIMI risk score for NSTEMI/UA categorises patients to be in low-risk at the score of <5 and the score of ≥5 to be in the high-risk category (Antman et al., 2000). As for the ML models, in this study, the low- and high-risk patients were categorised based on the study by Correia et al. (2014) and personal correspondence through email with the main author, Luis Correia from the study "Prognostic value of TIMI score versus GRACE score in ST-segment elevation myocardial infarction" (Appendix D). Their study stated that the rate of death exceeding 8% is considered high risk in patients' mortality prediction.

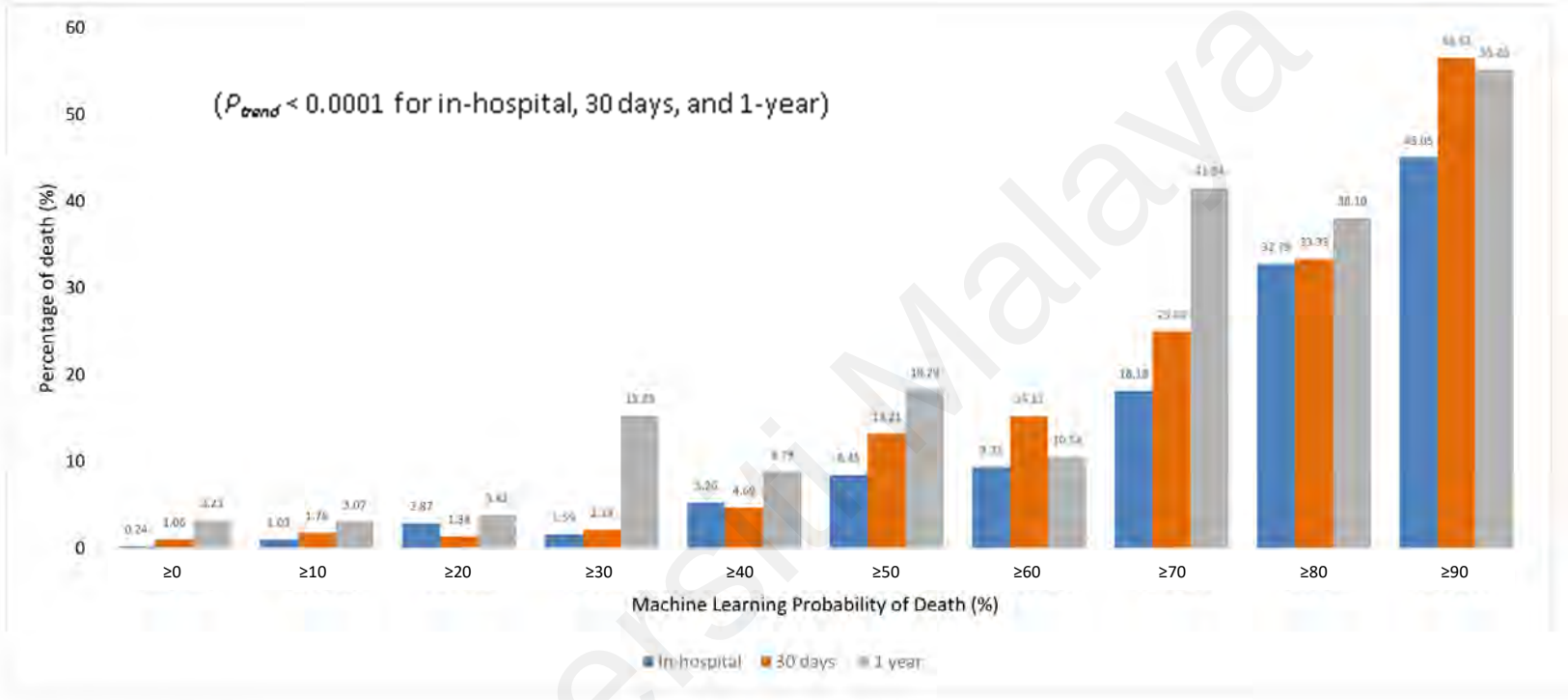
Hence, the cut-off points between the low- and the high-risk patients for all the best ML models were determined based on the rate of death graphs in Figures 4.4 and 4.6 below where the percentile that exceeded 8% rate of death is considered the cut-off point of the model and are shown in Figure 4.7 below, according to each of their time points and data population.



Total number of patients

In-hospital	41	182	388	353	280	283	136	125	53	48
30 days	10	101	190	154	155	136	65	71	20	36
1 year	10	94	188	135	135	134	75	55	21	33

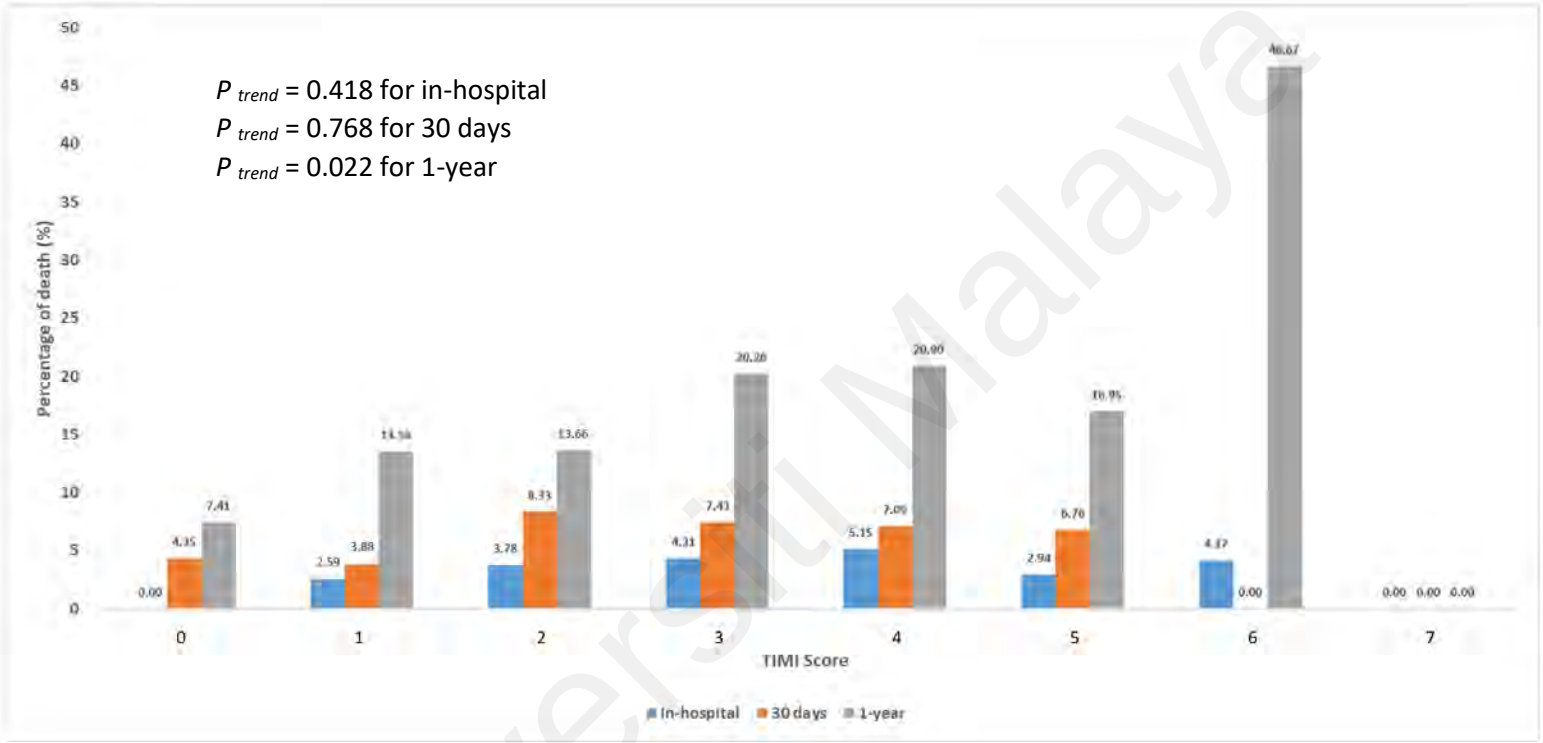
Figure 4.3: The rate of death across the TIMI risk score in the STEMI dataset.



Total number of patients

In-hospital	849	292	174	126	95	71	75	55	61	91
30 days	188	227	145	94	64	53	46	36	39	46
1 year	124	163	131	85	91	82	38	65	63	38

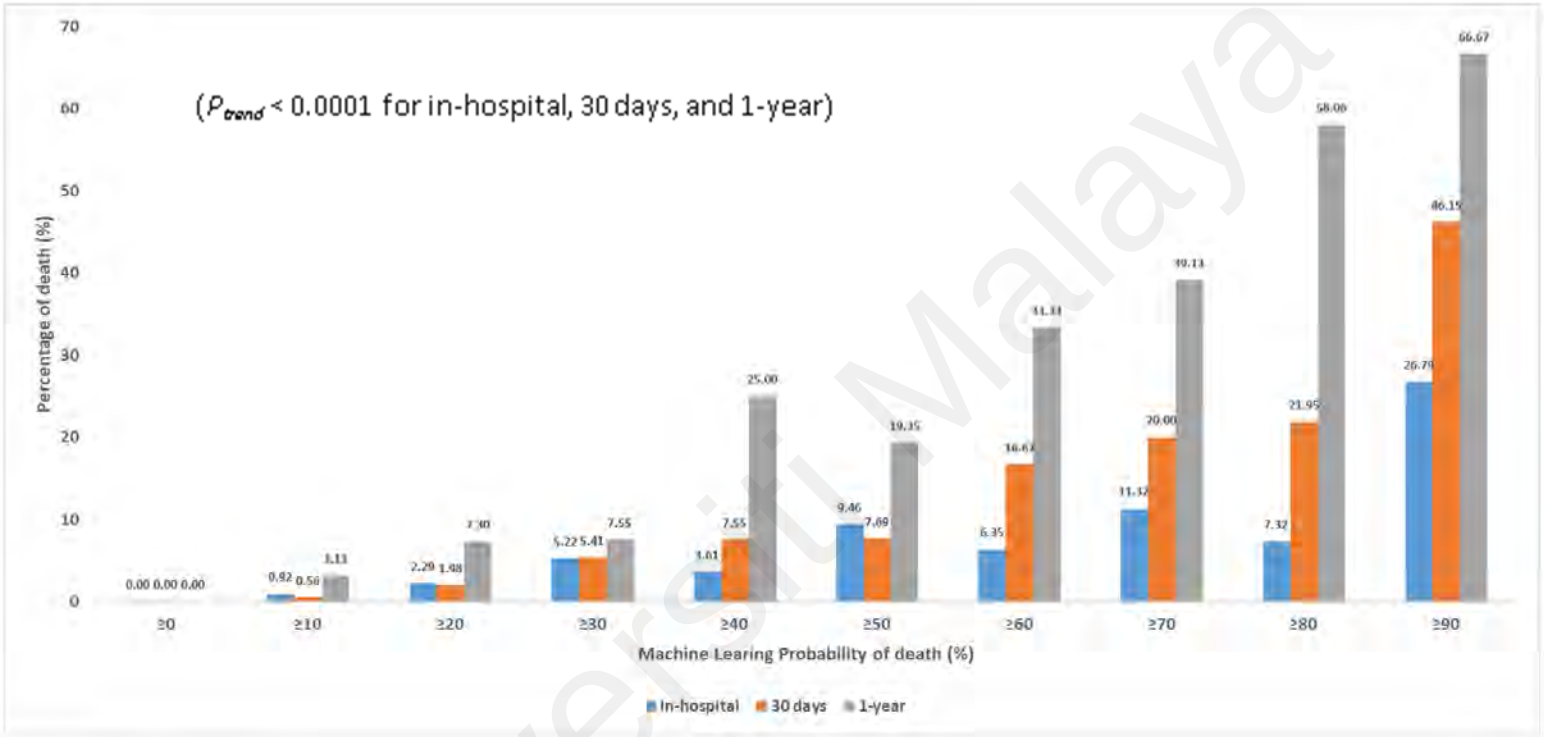
Figure 4.4: The rate of death across ML probability of death in STEMI dataset.



Total number of patients

In-hospital	104	232	370	325	272	102	24	2
30 days	23	103	192	175	141	74	12	0
1-year	27	118	183	153	134	59	15	2

Figure 4.5: The rate of death across the TIMI risk score in the NSTEMI/UA dataset.



Total number of patients

In-hospital	426	326	175	134	83	74	63	53	41	56
30 days	122	180	101	74	53	52	36	35	41	26
1-year	29	128	137	106	80	62	57	46	31	15

Figure 4.6: The rate of death across ML probability of death in NSTEMI/UA dataset.

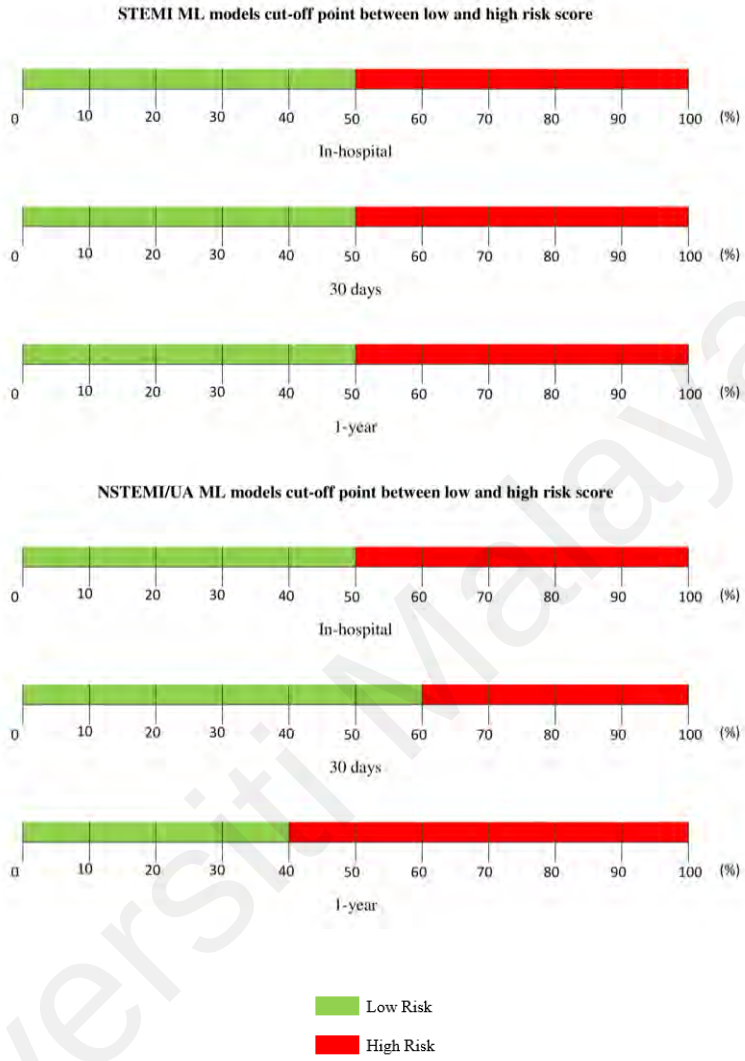


Figure 4.7: The cut-off point between the low and high-risk patients based on the 30% testing dataset ML best model performances.

If the model is performing well in predicting the mortality of patients, it will classify patients who are high-risk and most likely to die in the high-risk group and patients who are low-risk and less likely to die in the low-risk group. This information can be extracted from the graphs above to further compare the performance of the best ML models against the TIMI risk scores. In the high-risk group of the STEMI dataset, ML was better in predicting

mortality in comparison to TIMI risk score for in-hospital death (21.94% vs 16.15%) but almost similar for prediction for 30 days and 1-year deaths (25.61% vs 23.15% and 35.71% vs 34.48%, respectively). ML models predicted a better mortality rate when compared to TIMI risk score in the lower risk group for in-hospital (1.97% vs 2.59%), 30 days (1.73% vs 3.46%), and 1-year (5.05% vs 9.35%). This was almost similar to the NSTEMI/UA dataset where in the high-risk group, ML models predicted the mortality better as compared to the TIMI risk score in terms of percentage difference for in-hospital, 30 days, and 1-year (with the value of 12.20% vs 3.13%, 24.64% vs 5.81%, and 33.33% vs 22.37% respectively). As for the lower-risk patients, ML was still better as compared to the TIMI risk score in predicting the mortality based on the percentage of deaths in in-hospital (1.49% vs 3.68%), 30 days (2.58% vs 6.94%), and 1-year (5.50% vs 16.59%).

On the other hand, classification of patients based on the mortality prediction cut-off points above and categorising them between alive (below the cut-off point) and dead (above the cut-off point) were further discussed in the implementation of NRI. Through NRI, the sum of the net proportion of people with (dead) and without (alive) the event who were correctly assigned from the old/conventional prediction model (in this case, the TIMI risk score) and a new prediction model (the best ML prediction models) were distinguished.

In the STEMI dataset, the NRI for the in-hospital mortality prediction, the net reclassification of patients improved using the best ML model produced a net reclassification improvement of 0.20 with $p<0.0001$ over the original TIMI risk score, that is, a 20% improved classification. NRI for 30 days mortality prediction reported the net reclassification of patients improved using the best ML model produced a net reclassification improvement of 0.18 with $p<0.0001$ over the original TIMI risk score, that is, a 18% improved classification. In the 1-year mortality prediction, the net reclassification

of patients improved using the best ML model produced a net reclassification improvement of 0.14 with p<0.0001 over the original TIMI risk score, that is, a 14% improved classification (Table 4.11).

Table 4.11:Predicted risks and reclassification of STEMI patient's mortality between ML best model (SVMvarImp- SBE-SVM) and TIMI risk score for in-hospital, 30-days, and 1-year on 30% validation dataset.

<table>
<tr><td colspan="7">In-hospital</td></tr>
<tr><td colspan="7">Individuals with events (n=101)</td></tr>
<tr><td rowspan="5"></td><td rowspan="3"></td><td colspan="2" rowspan="2">Numbers of individuals</td><td colspan="2" rowspan="2">Reclassified</td><td rowspan="2">Net correctly reclassified (%)</td></tr>
<tr></tr>
<tr><td>Low risk</td><td>High risk</td><td>Increased risk</td><td>Decreased risk</td><td rowspan="4"></td></tr>
<tr><td colspan="2">Machine Learning</td><td rowspan="3">23</td><td rowspan="3">4</td></tr>
<tr><td>TIMI score</td><td colspan="2"></td></tr>
<tr><td>Low risk</td><td>13</td><td>23</td><td>18.81</td></tr>
<tr><td>High risk</td><td>4</td><td>61</td></tr>
<tr><td colspan="7">Individuals without events (n=1788)</td></tr>
<tr><td rowspan="4"></td><td></td><td colspan="2">Machine Learning</td><td rowspan="4">116</td><td rowspan="4">144</td><td rowspan="4">1.57</td></tr>
<tr><td>TIMI score</td><td colspan="2"></td></tr>
<tr><td>Low risk</td><td>1375</td><td>116</td></tr>
<tr><td>High risk</td><td>144</td><td>153</td></tr>
<tr><td>Net reclassification Index (NRI) (%)</td><td colspan="6">20.38</td></tr>
<tr><td>p-value</td><td colspan="6"><0.0001</td></tr>
<tr><td colspan="7">30 days</td></tr>
<tr><td colspan="7">Individuals with events (n=75)</td></tr>
<tr><td></td><td></td><td colspan="2">Number of individuals</td><td colspan="2">Reclassified</td><td>Net correctly reclassified (%)</td></tr>
</table>

Table 4.11, continued.

		Low risk	High risk	Increased risk	Decreased risk	
		Machine Learning				
	TIMI score					
	Low risk	11	17	17	2	20.00
	High risk	2	45			
Individuals without events (n=863)						
		Machine Learning				
	TIMI score					
	Low risk	652	66	66	53	-1.51
	High risk	53	92			
Net reclassification index (NRI) (%)						18.49
p-value						<0.0001
1-year						
Individuals with events (n=126)						
		Number of individuals		Reclassified		Net correctly reclassified (%)
		Low risk	High risk	Increased risk	Decreased risk	
		Machine Learning				
	TIMI score					
	Low risk	21	44	44	14	23.81
	High risk	14	47			
Individuals without events (n=754)						
		Machine Learning				
	TIMI score					
	Low risk	523	108	108	36	-9.55
	High risk	36	87			

Table 4.11, continued.

Net reclassification index (NRI) (%)	14.26
p-value	<0.0001

In the NSTEMI/UA dataset, the net reclassification of patients for in-hospital mortality prediction improved using the best ML model, yielding a net reclassification improvement of 0.56 with p<0.0001 over the original TIMI risk score, or a 56% improved classification. The net reclassification of patients increased using the best ML model for 30 days mortality prediction, with a net reclassification improvement of 0.56 and p<0.0001 over the initial TIMI risk score, indicating a 56% improvement in classification. The net reclassification of patients also increased using the ML in the 1-year model, with a net reclassification improvement of 0.44 with p<0.0001 over the original TIMI risk score, indicating a 44% improvement in classification (Table 4.12).

Table 4.12: Predicted risks and reclassification of NSTEMI/UA patient's mortality between ML best model (SVMvarImp- SBE-SVM) and TIMI risk score for in-hospital, 30-days, and 1-year on 30% validation dataset.

In-hospital						
Individuals with events (n=52)						
		Numbers of individuals		Reclassified		Net correctly reclassified (%)
		Low risk	High risk	Increased risk	Decreased risk	
		Machine Learning				
	TIMI score					
	Low risk	16	35	35	1	65.38
	High risk	1	0			

Table 4.12, continued.

Individuals without events (n=1379)						
		Machine Learning				
	TIMI score					
	Low risk	1038	217	217	89	-9.28
	High risk	89	35			
Net reclassification Index (NRI) (%)						56.10
p-value						<0.0001
30 days						
Individuals with events (n=49)						
		Number of individuals		Reclassified		Net correctly reclassified (%)
		Low risk	High risk	Increased risk	Decreased risk	
		Machine Learning				
	TIMI score					
	Low risk	14	30	30	1	59.18
	High risk	1	4			
Individuals without events (n=671)						
		Machine Learning				
	TIMI score					
	Low risk	506	84	84	61	-3.43
	High risk	61	20			
Net reclassification index (NRI) (%)						55.76
p-value						<0.0001
1-year						
Individuals with events (n=119)						

Table 4.12, continued.

		Number of individuals		Reclassified		Net correctly reclassified (%)
		Low risk	High risk	Increased risk	Decreased risk	
		Machine Learning				
	TIMI score					
	Low risk	21	81	81	1	67.23
	High risk	1	16			
Individuals without events (n=572)						
		Machine Learning				
	TIMI score					
	Low risk	351	162	162	27	-23.60
	High risk	27	32			
Net reclassification index (NRI) (%)						43.63
p-value						<0.0001

4.5 DL performance

As noted in Chapter 3, the development of the DL model for mortality prediction in this study is an extension of the ML study's main future compass. In general, the same steps of performance evaluation were implemented from ML prediction models to the DL prediction models to see how well DL might work with this type of data. The performance of the DL algorithms in predicting mortality was tested using all the variables and the selected variables from the best ML models for both STEMI and NSTEMI/UA datasets as DL cannot perform feature selection. In this case, SVMvarImp-SBE-SVM selected variables from the best ML model were implemented into the DL models for in-hospital, 30

days and 1-year mortality prediction. The maximal performance of the DL models with all and selected variables using the 30% untouched testing data is tabulated in Table 4.13 below. Table 4.14 provides a detailed performance evaluation of the best DL model for all three time points.

Table 4.13: The AUC of DL models with all variables and selected variables from the best ML model based on a 30% testing dataset.

DL model	The area under the ROC Curve (95% CI)		
	In-hospital	30 days	1-year
STEMI			
All	0.97 (0.9625-0.9729)	0.96 (0.9500-0.9665)	0.90 (0.8814-0.9114)
Selected	0.96 (0.9524-0.9632)	0.93 (0.9241-0.9443)	0.90 (0.8814-0.9097)
NSTEMI/UA			
All	0.98 (0.9718-0.9824)	0.93 (0.9217-0.9461)	0.85 (0.8294-0.8695)
Selected	0.97 (0.9649-0.9745)	0.91 (0.8916-0.9200)	0.88 (0.8614-0.8963)

Table 4.14: Additional performance metrics based on 30% testing dataset for DL models with all variables and selected variables from the best ML model.

	PPV	NPV	Sensitivity	Specificity	Accuracy (CI 95%)	Mcnemar's test (p-value)
STEMI						
In-hospital						
All	0.276	0.969	0.475	0.930	0.905 (0.891,0.918)	7.385e-08
Selected	0.251	0.981	0.693	0.883	0.873 (0.857,0.888)	<2e-16
30 days						
All	0.368	0.957	0.520	0.922	0.890 (0.868,0.910)	0.003117
Selected	0.301	0.978	0.787	0.841	0.837 (0.812,0.860)	<2e-16

Table 4.14, continued.

	PPV	NPV	Sensitivity	Specificity	Accuracy (CI 95%)	Mcnemar's test (p-value)
1 year						
All	0.432	0.919	0.532	0.883	0.833 (0.807,0.857)	0.02092
Selected	0.270	0.968	0.881	0.602	0.642 (0.609,0.674)	<2e-16
NSTEMI/UA						
In-hospital						
All	0.224	0.973	0.288	0.962	0.938 (0.924,0.950)	0.1378
Selected	0.173	0.983	0.577	0.896	0.885 (0.867,0.901)	<2e-16
30 days						
All	0.159	0.945	0.286	0.890	0.849 (0.820,0.874)	2.73e-04
Selected	0.195	0.981	0.796	0.760	0.763 (0.730,0.793)	<2e-16
1 year						
All	0.375	0.885	0.479	0.834	0.773 (0.740,0.804)	0.01065
Selected	0.241	0.945	0.882	0.423	0.502 (0.464,0.540)	<2e-16

From the Table 4.13 above, the AUC value between the models with all variables and selected variables was almost the same. In the STEMI dataset model performance, all the models with selected variables have lower AUC values as compared to the models with all variables except for the 1-year mortality prediction where the AUC value was similar. DL models with all variables performed slightly better than the models with selected variables for in-hospital (AUC = 0.97, 95% CI: 0.963 to 0.973, vs AUC = 0.96, 95% CI: 0.953 to 0.963) and 30 days (AUC = 0.96, 95% CI: 0.950 to 0.967, vs AUC = 0.93, 95% CI: 0.924

to 0.944) while the performance of 1-year mortality prediction between the model with all and selected variable was similar (AUC = 0.90, 95% CI: 0.811 to 0.91, vs AUC = 0.90, 95% CI:0.881 to 0.910). The comparison between the models (all and selected variables) were also significant ($p<0.0001$) in the three time points; in-hospital (95% CI: 0.064 to 0.083), 30 days (95% CI: 0.111 to 0.141) and 1-year (95% CI: 0.105 to 0.137).

For the NSTEMI/UA dataset, the models with all variables performed significantly better ($p<0.0001$) than the models with the selected variables for in-hospital (95% CI, 0.080 to 0.103) and 30 days (95% CI, 0.114 to 0.156) with the AUC of 0.98, 95% CI: 0.972 to 0.982 vs 0.97, 95% CI: 0.965 to 0.975, and AUC of 0.93, 95% CI: 0.922 to 0.946 vs 0.91, 95% CI: 0.892 to 0.920 respectively. Anyhow, for the 1-year mortality prediction, the performance of the model with selected variables was significantly better ($p<0.0001$; 95% CI: 0.054 to 0.089) as compared to the model with all variables (AUC = 0.88, 95% CI: 0.861 to 0.896 vs AUC = 0.85, 95% CI: 0.829 to 0.870).

Consequently, based on the result above, for both dataset STEMI and NSTEMI/UA, models with the selected variables were considered as the best model due to their lower number of variables in the models and the difference in AUC values between the models (all vs selected variables) were very small.

For the secondary analysis, similar to the ML, missing values in each variable were imputed and utilised as the training dataset for model development, with the 30% untouched testing dataset used for model performance evaluation. Secondary analysis of the best DL models (DL models with selected variables) performance trained with imputed data for the STEMI dataset reported an AUC of 0.95 (95% CI: 0.942 to 0.955) for in-hospital, 0.93 (95% CI: 0.918 to 0.940) for 30 days and 0.90 (95% CI: 0.890 to 0.917) for 1-year mortality predictions. The AUC values of the imputed dataset and the complete

dataset were almost similar and highly significant for in-hospital, 30 days and 1-year ($p<0.0001$; 95% CI: 0.028 to 0.040, 95% CI: 0.016 to 0.036, 95% CI: 0.093 to 0.110 respectively).

The secondary analysis of the best DL models for the NSTEMI/UA dataset reported an AUC of 0.95 (95% CI: 0.945-0.960) for in-hospital mortality prediction, 0.90 (95% CI: 0.881-0.911) for 30 days, and 0.86 (95% CI: 0.837-0.873) for 1-year mortality prediction. DL models with complete cases produced very similar AUC results and were highly significant for in-hospital ($p<0.0001$,95% CI: 0.037 to 0.040), 30 days ($p<0.0001$,95% CI: 0.020 to 0.039), and 1-year ($p=0.0001$, 95% CI:0.007 to 0.028)

The graphs in Figure 4.8 depict the DL best model performances against TIMI risk score and ML best models for both STEMI and NSTEMI/UA datasets. The best-selected models for DL performed better against the ML best models (SVMvarImp-SBE-SVM) based on the AUC value using the 30% untouched dataset. In the STEMI dataset, there was a significant difference ($p<0.0001$) in comparison between the best models from DL and ML for in-hospital (95% CI: 0.073 to 0.083), 30 days (95% CI: 0.096 to 0.114) and 1-year (95% CI: 0.071 to 0.085). Comparison between the best DL and ML models for the NSTEMI/UA dataset also reported a significant difference ($p<0.0001$) for all three time points (95% CI: 0.140 to 0.158 for in-hospital, 95% CI: 0.073 to 0.091 for 30 days and 95% CI: 0.073 to 0.088 for 1-year mortality prediction).

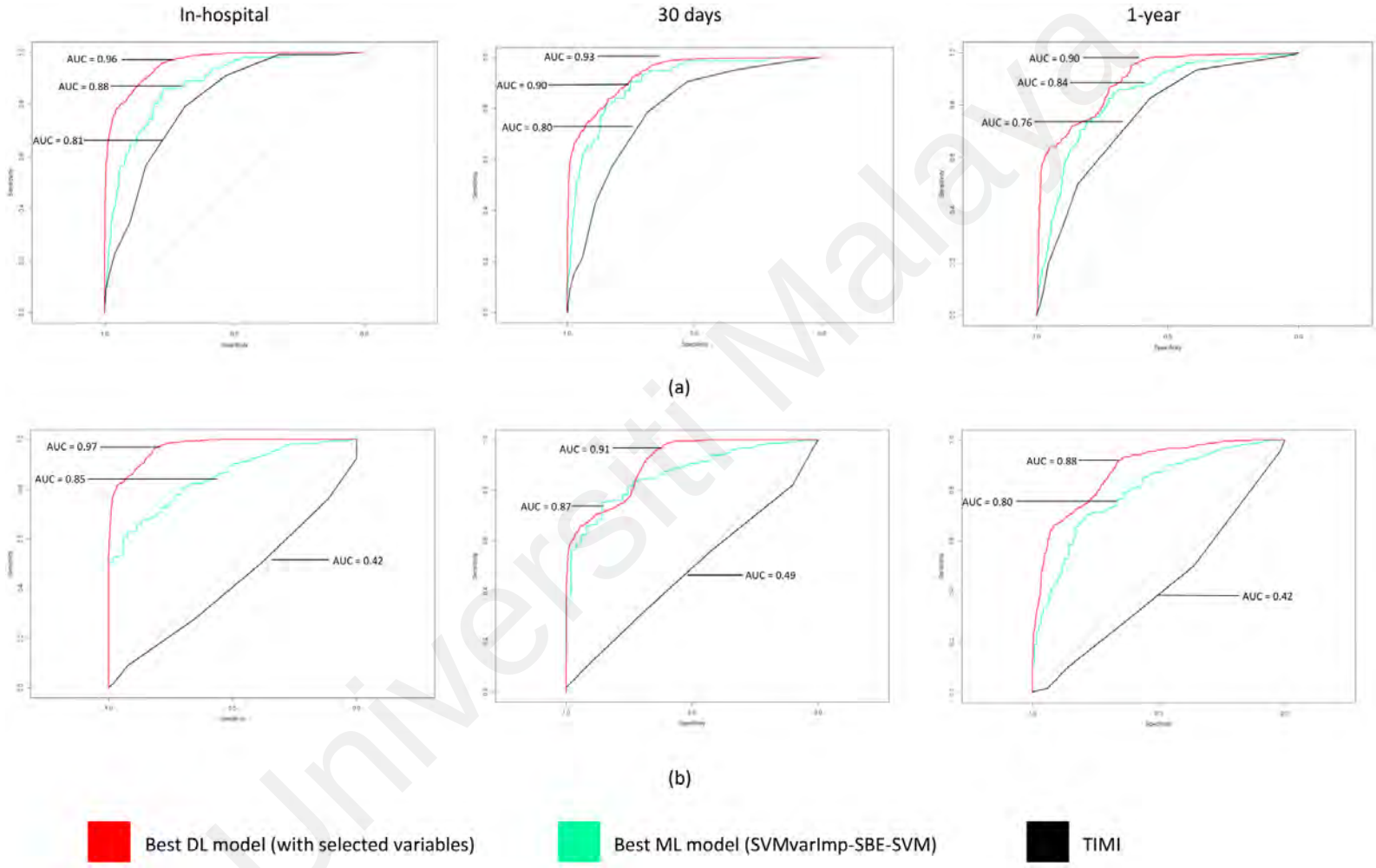
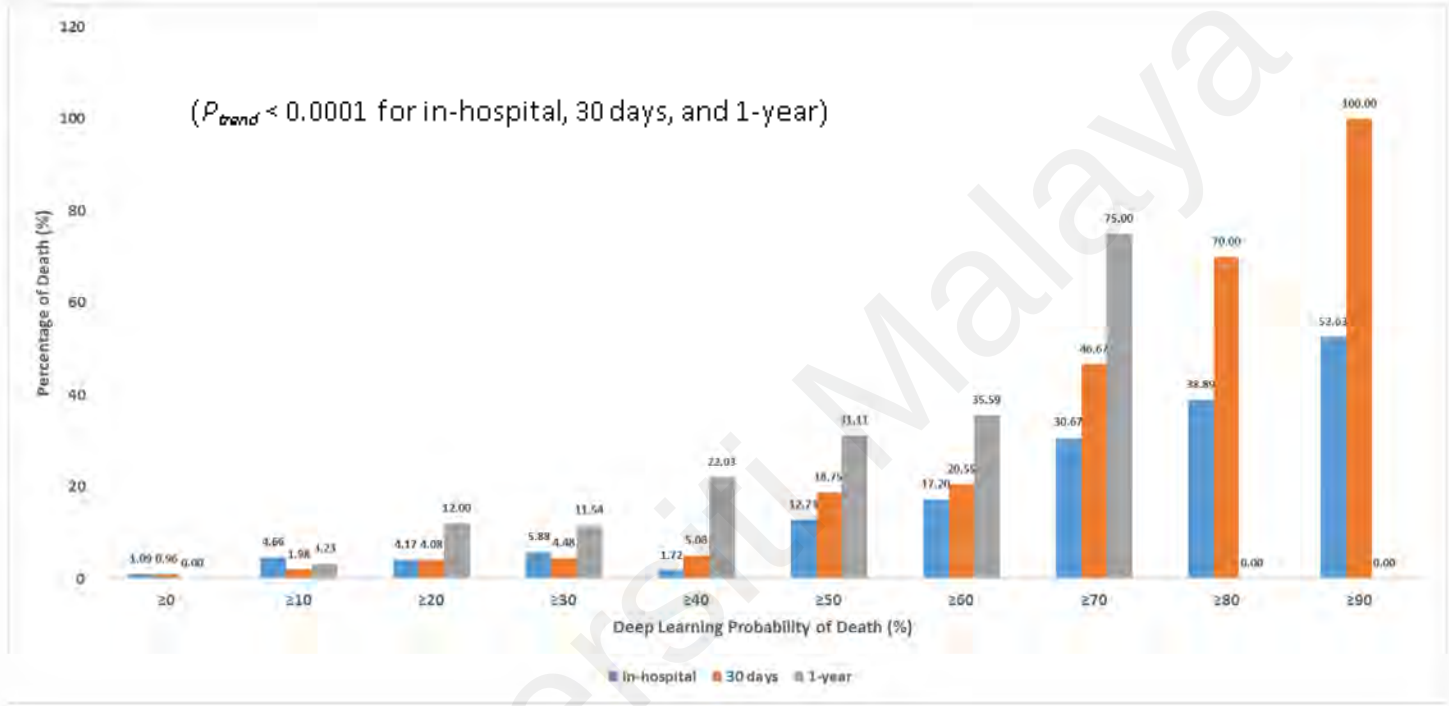


Figure 4.8: The receiver operating characteristic (ROC) curves of DL best model, ML best model, and TIMI score based on a 30% testing dataset. (a) ROC curves show the performance of in-hospital, 30 days, and 1-year mortality prediction of the STEMI dataset. (b) ROC curves show the performance of in-hospital, 30 days, and 1-year mortality prediction of the NSTEMI/UA dataset.

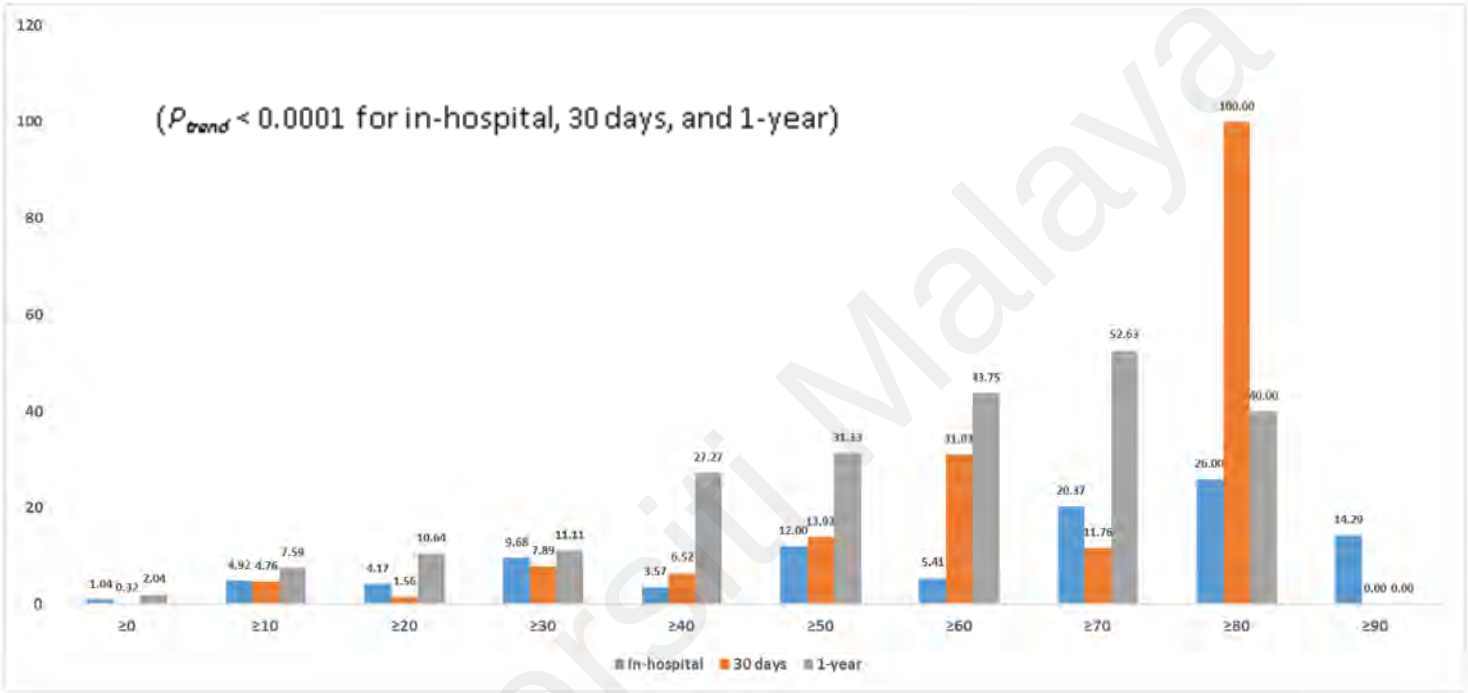
As can be seen from the graphs above, the best-selected DL models in both STEMI and NSTEMI/UA datasets also performed better against TIMI based on the AUC value using the untouched 30% validation dataset ($p < 0.0001$ for all models). Similar to the ML best model's analysis, the mortality rate of the DL best models was also produced and compared in Figure 4.9 (for STEMI) and 4.10 (for NSTEMI/UA) below. Using the same 8% value of the rate of death from the Correia et al. (2014) study, the cut-off points between the low- and high-risk patients were determined and depicted in Figure 4.11 according to their time points and dataset.



Total number of patients

In-hospital	1196	193	96	68	58	55	93	75	36	19
30 days	417	101	98	67	59	64	73	45	10	4
1-year	4	465	75	26	59	180	59	12	0	0

Figure 4.9: The rate of death across DL probability of death in STEMI dataset.



Total number of patients

In-hospital	1053	122	24	31	28	25	37	54	50	7
30 days	311	63	64	38	46	122	58	17	1	0
1-year	98	158	188	27	33	83	80	19	5	0

Figure 4.10: The rate of death across DL probability of death in NSTEMI/UA dataset.

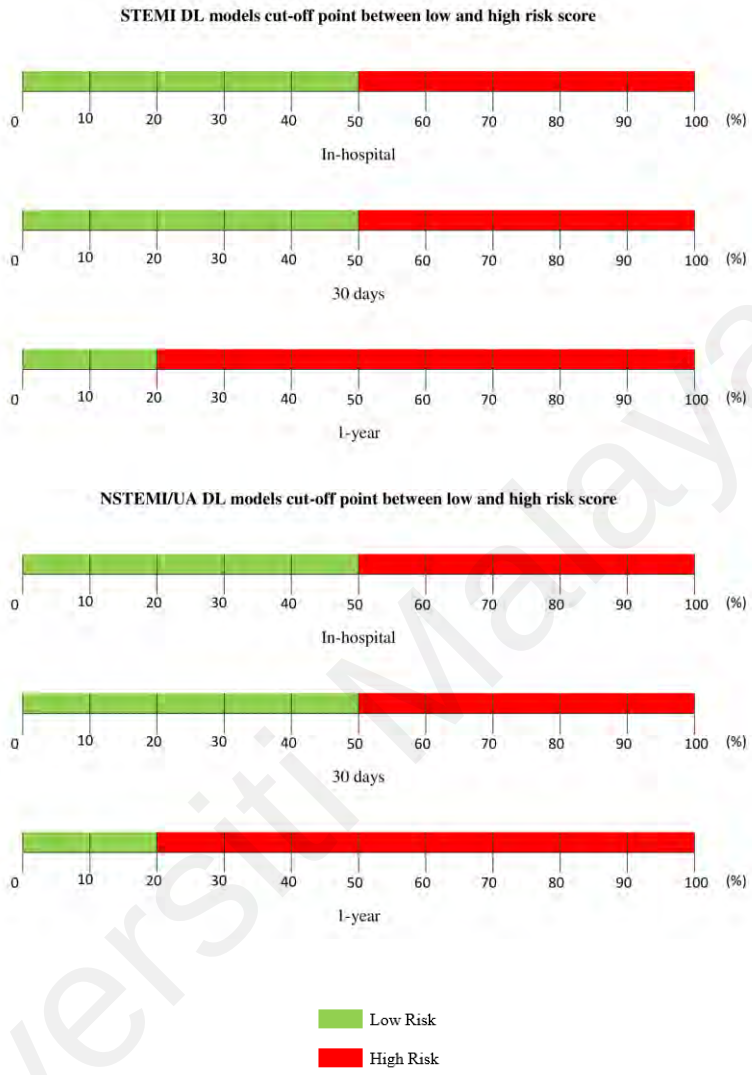


Figure 4.11: The cutoff point between the low and high-risk patients based on the 30% testing dataset DL best model performances.

In the high-risk group of the STEMI dataset, DL best models predicted better mortality in comparison to TIMI for in-hospital (25.19% vs 16.15%), 30 days (30.13% vs 23.15%) but similar in 1-year (both with 27%). DL best models predicted a better mortality rate when compared to the TIMI risk score in the low-risk group with lower numbers of death

for in-hospital (1.92% vs 2.59%), 30 days (2.16% vs 3.46%), and 1 year (3.20% vs 9.34%). The same pattern can be seen in the NSTEMI/UA dataset, where the best DL models were able to predict better mortality in the high-risk group than the TIMI risk score for all three time points (In-hospital, 17.3% vs 3.13%; 30 days, 19.19% vs 5.81%; 1-year, 24.14% vs 22.36%). As for the low-risk group, DL still predicted an almost better mortality rate than the TIMI risk score for in-hospital (1.74% vs 3.68%), but better prediction on 30 days (2.11% vs 6.94%) and 1-year (5.47% vs 16.59%) mortality prediction.

Regarding the NRI of the STEMI dataset for the best in-hospital mortality prediction, the net reclassification of patients improved using the DL best model producing a net reclassification improvement of 0.14 with $p<0.0001$ over the TIMI risk score, that is, a 14% improved classification. The net reclassification of patients improved using the DL best model for 30 days mortality prediction, yielding a net reclassification improvement of 0.15 with $p<0.0001$ over the initial TIMI risk score, or a 15% improvement in classification. The net reclassification of patients increased by 16% improvement in classification using the ML in the 1-year model, with a net reclassification improvement of 0.16 with $p<0.0001$ over the original TIMI risk score (Table 4.15).

Table 4.15: Predicted risks and reclassification of STEMI patient's mortality between DL best model (with selected variables) and TIMI risk score for in-hospital, 30-days, and 1-year on 30% validation dataset.

In-hospital						
Individuals with events (n=101)						
		Numbers of individuals		Reclassified		Net correctly reclassified (%)
		Low risk	High risk	Increased risk	Decreased risk	

Table 4.15, continued.

		Machine Learning				
	TIMI score					
	Low risk	19	20	20	12	7.92
	High risk	12	50			
Individuals without events (n=1788)						
		Machine Learning				
	TIMI score					
	Low risk	1378	88	88	202	6.38
	High risk	202	120			
Net reclassification Index (NRI) (%)						14.30
p-value						<0.0001
30 days						
Individuals with events (n=75)						
		Number of individuals		Reclassified		Net correctly reclassified (%)
		Low risk	High risk	Increased risk	Decreased risk	
		Machine Learning				
	TIMI score					
	Low risk	10	15	15	6	12.00
	High risk	6	44			
Individuals without events (n=863)						
		Machine Learning				
	TIMI score					
	Low risk	644	53	53	82	3.36
	High risk	82	84			
Net reclassification index (NRI) (%)						15.36

Table 4.15, continued.

p-value						<0.0001
1-year						
Individuals with events (n=126)						
		Number of individuals		Reclassified		Net correctly reclassified (%)
		Low risk	High risk	Increased risk	Decreased risk	
		Machine Learning				
	TIMI score					
	Low risk	12	54	54	3	40.48
	High risk	3	57			
Individuals without events (n=754)						
		Machine Learning				
	TIMI score					
	Low risk	432	208	208	22	-24.76
	High risk	22	92			
Net reclassification index (NRI) (%)						15.81
p-value						<0.0001

In the NSTEMI/UA dataset, the NRI for the in-hospital mortality prediction, the net reclassification of patients improved using the DL best model produced a net reclassification improvement of 0.49 with $p<0.0001$ over the original TIMI risk score, that is, a 49% improved classification. The DL best model for 30 days mortality prediction improved the net reclassification of patients, with a net reclassification improvement of 0.56 and $p<0.0001$ compared to the baseline TIMI risk score, suggesting a 56% improvement in classification. The net reclassification of patients increased using the DL

best model in the 1-year mortality prediction, with a net reclassification improvement of 0.27 with p<0.0001 over the original TIMI risk score, indicating a 27% improvement in classification (Table 4.16).

Table 4.16: Predicted risks and reclassification of NSTEMI/UA patient's mortality between DL best model (with selected variables) and TIMI risk score for in-hospital, 30-days, and 1-year on 30% validation dataset.

<table>
<tr><td colspan="7">In-hospital</td></tr>
<tr><td colspan="7">Individuals with events (n=52)</td></tr>
<tr><td rowspan="5"></td><td rowspan="3"></td><td colspan="2" rowspan="2">Numbers of individuals</td><td colspan="2" rowspan="2">Reclassified</td><td rowspan="2">Net correctly reclassified (%)</td></tr>
<tr></tr>
<tr><td>Low risk</td><td>High risk</td><td>Increased risk</td><td>Decreased risk</td><td></td></tr>
<tr><td colspan="3">Machine Learning</td><td rowspan="4">26</td><td rowspan="4">0</td><td rowspan="4">50.00</td></tr>
<tr><td>TIMI score</td><td colspan="2"></td></tr>
<tr><td>Low risk</td><td>22</td><td>26</td></tr>
<tr><td>High risk</td><td>0</td><td>4</td></tr>
<tr><td colspan="7">Individuals without events (n=1379)</td></tr>
<tr><td rowspan="4"></td><td></td><td colspan="2">Machine Learning</td><td rowspan="4">126</td><td rowspan="4">107</td><td rowspan="4">-1.38</td></tr>
<tr><td>TIMI score</td><td colspan="2"></td></tr>
<tr><td>Low risk</td><td>1129</td><td>126</td></tr>
<tr><td>High risk</td><td>107</td><td>17</td></tr>
<tr><td>Net reclassification Index (NRI) (%)</td><td colspan="6">48.62</td></tr>
<tr><td>p-value</td><td colspan="6"><0.0001</td></tr>
<tr><td colspan="7">30 days</td></tr>
<tr><td colspan="7">Individuals with events (n=49)</td></tr>
<tr><td rowspan="2"></td><td rowspan="2"></td><td colspan="2" rowspan="2">Number of individuals</td><td colspan="2" rowspan="2">Reclassified</td><td rowspan="2">Net correctly reclassified (%)</td></tr>
<tr></tr>
<tr><td>Low risk</td><td>High risk</td><td>Increased risk</td><td>Decreased risk</td><td></td></tr>
</table>

Table 4.16, continued.

		Machine Learning				
	TIMI score					
	Low risk	11	33	33	0	67.35
	High risk	0	5			
Individuals without events (n=671)						
		Machine Learning				
	TIMI score					
	Low risk	464	126	126	47	-11.77
	High risk	47	34			
Net reclassification index (NRI) (%)						55.58
p-value						<0.0001
1-year						
Individuals with events (n=119)						
		Number of individuals		Reclassified		Net correctly reclassified (%)
		Low risk	High risk	Increased risk	Decreased risk	
		Machine Learning				
	TIMI score					
	Low risk	13	89	89	1	73.95
	High risk	1	16			
Individuals without events (n=572)						
		Machine Learning				
	TIMI score					
	Low risk	230	283	283	12	-47.38
	High risk	12	47			
Net reclassification index (NRI) (%)						26.57

Table 4.16, continued.

p-value	<0.0001

4.6 Application of the algorithms in clinical practice

Cardiac catheterization is the most often used early therapy for STEMI and frequently NSTEMI patients. It is only available in hospitals with cardiac services on-site. In Malaysia, these may be available in private, university, or private hospitals. There are 10 cardiac catheterization centres under the Ministry of Health across the country and 3 university hospitals participating in the NCVD Registry. They receive referrals for cardiac cases from across the country and prioritising access is of utmost importance. As patients with ACS, maybe from within or without these cardiac centres, it is important for patients with the best likelihood to improve with cardiac catheterization to be selected and offered the intervention.

Having cardiac catheterization performed on patients means patients will be better risk-stratified, as their anatomy is delineated, and their ischaemic and non-ischaemic risks are addressed. Patients may then be offered ongoing and optimisation of medical therapy or revascularisation options, whether with angioplasty or bypass surgery. This could explain why early catheterization has been proven to improve mortality in available literature such as in the DANAMI-2 trial (Andersen, 2003; Busk, 2008; Thrane, 2020).

Access to cardiac catheterization in Malaysia is limited due to the volumes of patients referred as well as the cost related to performing cardiac catheterization. In most public cardiac centres, waiting time for this procedure may take months before it is made

available. Identifying patients with ACS who may benefit from a cardiac catheterization will help improve access as well as reduce the cost burden to the government and patients.

The algorithms used in this study will generate new probabilities of death for the patient based on whether or not the patient receives cardiac catheterization. Ultimately, the (increasing or decreasing) probability of death of the algorithms will assist the medical practitioner whether or not the patient should proceed with the therapy, hence referred for cardiac catheterization early and subsequently having better outcomes with reduced likelihood of death.

In order to test out the applicability of the algorithms:

1. The input of those patients who did not have a cardiac catheterization was changed to have undergone cardiac catheterization.
2. The input of those patients who had have a cardiac catheterization was changed to not having one.

In the first situation above, this study wanted to see how many patients were switched from high-risk to low-risk once cardiac catheterization input was altered. Additionally, this study calculated the change in the percentage of probability of death between before and after the input was adjusted by more than 10% (considered excessive). The likelihood of the patients dying in this circumstance should, theoretically, be lower than the prior prediction. This is because cardiac catheterization should reduce the chance of a patient's death following therapy.

In the second scenario, this study intended to compare the patients' probability of death before and after changing the cardiac catheterization input and compute the percentage difference of the patients' mortality more than 10% (considered excessive). This is to

ensure that the algorithms did not over-predict the patients' outcomes following the change. Since the patient did not undergo cardiac catheterization, which was expected to minimise the patients' mortality, the likelihood of death should increase in this case.

This analysis was conducted using both ML and DL techniques on the STEMI and NSTEMI/UA datasets. However, for the STEMI dataset, this analysis was applied to in-hospital and 30-day mortality prediction, and for the NSTEMI/UA dataset, it was applied to 30-day and 1-year mortality prediction. This is because the cardiac catheterization variable was chosen based on the two independent datasets' predicted time points. Table 4.17 below summarises the result of the analysis and Figure 4.12 and Figure 4.13 show the probability of death correlation before and after the alteration of the cardiac catheterization input.

Table 4.17: The percentage change in the analysis of ML and DL algorithms used to predict the mortality of STEMI and NSTEMI/UA patients in opting for cardiac catheterization therapy.

STEMI			
Machine Learning			
	Without cardiac catheterization to with cardiac catheterization		With cardiac catheterization to without cardiac catheterization
	High Risk to Low Risk (%)	Difference more than 10% (%)	Difference more than 10% (%)
In-hospital	18.97*	26.67*	15.32*
30 days	17.05*	7.69*	0.41**

Table 4.17, continued.

Deep Learning			
	Without cardiac catheterization to with cardiac catheterization		With cardiac catheterization to without cardiac catheterization
	High Risk to Low Risk (%)	Difference more than 10% (%)	Difference more than 10% (%)
In-hospital	29.31*	37.36*	10.58*
30 days	27.00*	32.00*	9.38*
NSTEMI/UA			
Machine Learning			
	Without cardiac catheterization to with cardiac catheterization		With cardiac catheterization to without cardiac catheterization
	High Risk to Low Risk (%)	Difference more than 10% (%)	Difference more than 10% (%)
30 days	16.10*	9.32*	5.00*
1-year	5.15*	0.00*	0.00*
Deep Learning			
	Without cardiac catheterization to with cardiac catheterization		With cardiac catheterization to without cardiac catheterization
	High Risk to Low Risk (%)	Difference more than 10% (%)	Difference more than 10% (%)
30 days	24.26*	20.71*	0.00*

Table 4.17, continued

	Without cardiac catheterization to with cardiac catheterization		With cardiac catheterization to without cardiac catheterization
	High Risk to Low Risk (%)	Difference more than 10% (%)	Difference more than 10% (%)
1-year	10.51*	5.97*	2.41**

*p-value < 0.0001

**p-value > 0.05

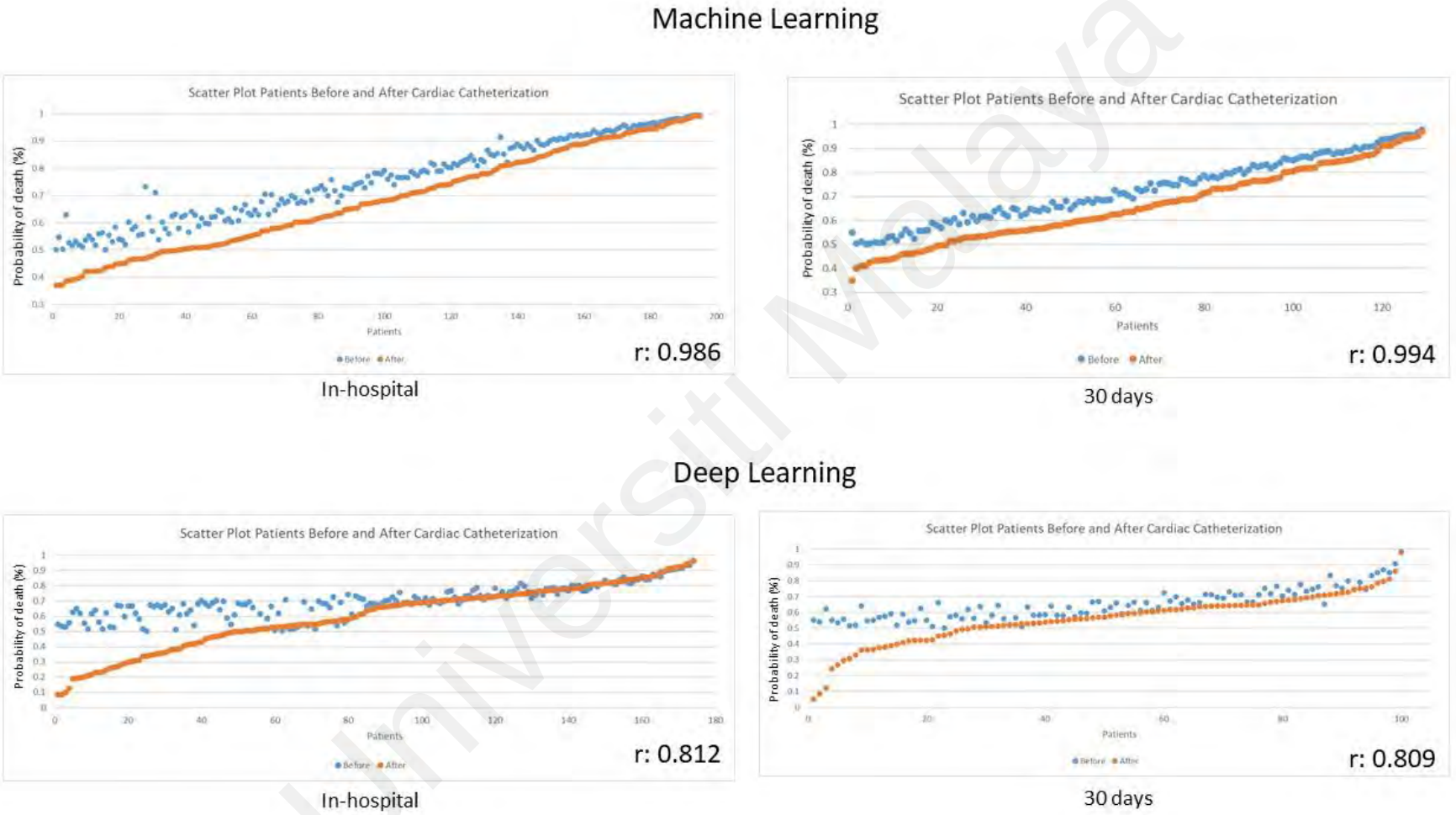


Figure 4.12: Plotting of the probability of death before and after the alteration of the cardiac catheterization input for STEMI dataset.

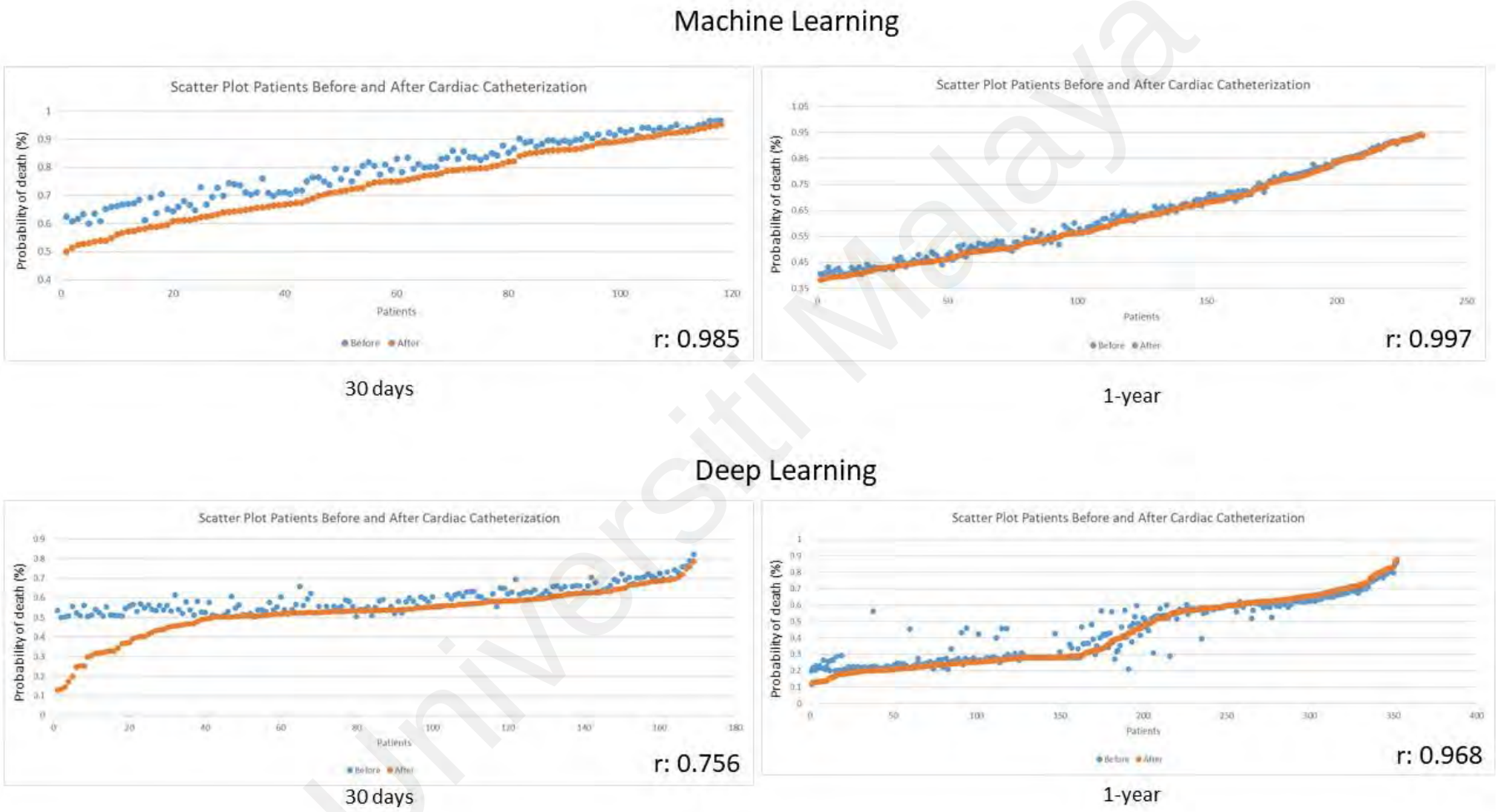


Figure 4.13: Plotting of the probability of death before and after the alteration of the cardiac catheterization input for NSTEMI/UA dataset.

Both ML and DL algorithms were used to see if cardiac catheterization can reduce risk mortality, and according to Table 4.17 above, for the STEMI dataset, 17–29% of high-risk patients who would improve their risk of mortality (decrease in likelihood) were identified if they had received cardiac catheterization treatment and were re-classified from high risk to lower risk. In 7-37% of cases, offering cardiac catheterization to high-risk patients reduced mortality by more than 10%. ML and DL algorithms, on the other hand, were able to show that patients who were meant to have cardiac catheterization but did not get cardiac catheterization had a mortality increase of more than 10% in up to 15% of high-risk patients.

In the NSTEMI/UA dataset, using both ML and DL algorithms, this study found 5–24% of high-risk patients who might benefit from reduced risk mortality if they underwent cardiac catheterization therapy and were reclassified from high risk to lower risk. Offering cardiac catheterization to high-risk patients decreased mortality by more than 10% in up to 20% of patients. On the other hand, ML and DL algorithms were able to demonstrate that patients who were set for cardiac catheterization treatment but did not get it had a mortality increase of more than 10% in up to 5% of high-risk patients. This may be explained by the risk profile of this group of patients that is they were of high-risk (perhaps cardiogenic shock, renal failure, or multiple comorbidities) nature to start and submitting them to cardiac catheterization may result in further complications and therefore increased their risk of death.

Both Figures 4.12 and 4.13 show the correlation drawn between the probability of death before to and after cardiac catheterization treatment in high-risk patients utilising both ML and DL methods. All of the presented graphs show a high positive correlation, $r > 0.70$ ($p<0.0001$). As the graphs demonstrate, the majority of patients benefit from cardiac

catheterization and just a few have an increased mortality risk as a result of cardiac catheterization. These are rare circumstances in which the patient's health is likely to be severe and high risk, and a procedure like cardiac catheterization will only aggravate the patient's condition, increasing the chance of death.

In conclusion, this study has demonstrated that both the ML and DL algorithms were useful in identifying patients who would benefit or not benefit from cardiac catheterization therapy when they were presented with STEMI and NSTEMI/UA and whose outcome is validated at admission and 30 days for STEMI patients and 30 days and 1-year for NSTEMI/UA patients.

CHAPTER 5: DISCUSSION

Many risk scores and prediction models have been developed to this end using traditional statistical approaches that involve the inclusion of small, single-center cohorts, the application of numerous exclusion criteria, the inclusion of predefined variables expected to be related to the outcome, and the exclusion of variables that are not expected to be related to the outcome. These approaches do not take into consideration the potential prognostic value of interactions between several unexpectedly weaker risk factors and the primary outcome (Al'Aref et al., 2019). This study intended to leverage the power of big data analytics and AI algorithms (DL and ML) in order to construct an AI-based prediction model for the occurrence of Asian STEMI and NSTEMI/UA patients for in-hospital, 30-day, and 1-year mortality prediction.

In this study, mortality prediction models were developed using several ML algorithms (RF, SVM, and LR) in a multi-ethnic Asian patient in Malaysia. Their performances were comparable in predicting the short- and long-term mortality of patients with STEMI and NSTEMI/UA with those of traditional risk stratification. ML models performed better than TIMI for in-hospital, 30 days and 1-year (with AUC of 0.88 vs 0.81, 0.90 vs 0.80, 0.84 vs 0.76) for STEMI dataset and (AUC of 0.85 vs 0.42, 0.87 vs 0.49, 0.80 vs 0.42) for NSTEMI/UA dataset respectively. SVMvarImp-SBE-SVM for in-hospital, 30 days and 1-year mortality prediction had better performance compared to RF, LR, and TIMI scoring as well. Furthermore, adding more clinical variables to the models did enhance the performance of the predictive models for mortality in ACS. This is the distinctive feature of ML: it incorporates all patients and a considerable portion of variables (after excluding irrelevant or redundant variables) without requiring the use of significant exclusion criteria. The use of this methodology to clinical research could contribute to the development of

generally applicable prediction models and an improvement in the ability to predict future events, which is a primary focus in an era of precision medicine. Additionally, the ability to incorporate a large number of factors agnostically, without prior assumptions about likely significant predictors, eventually aided in the discovery of unique connections between specific traits and a particular outcome of interest.

5.1 ML model performance evaluation

For both STEMI and NSTEMI/UA datasets, ML algorithms outperformed the TIMI risk score with the best models (SVMvarImp-SBE-SVM) incorporating SBE as the feature selection methods, resulting in a smaller number of variables. ML algorithms have been suggested in some research as a way to improve the performance of the prognostic prediction model for patients with ACS both in the short and long term. In a recent study, AUC values of 0.84, 0.83, 0.84, and 0.85 were obtained in LR, NB, Cox regression, and RF, respectively, using only selected features, which outperformed the TIMI risk score (AUC: 0.77) in predicting STEMI in-hospital mortality (Li et al., 2017). Another study found that the AUCs in the best ML models were enhanced by 0.02 to 0.09 compared to that in the TIMI risk score when predicting patients with STEMI mortality outcomes after a 30-day follow-up (Shouval et al., 2017). According to Lee et al. (2021), the best selected ML models outperformed the traditional TIMI risk factor in predicting death for STEMI patients in-hospital (AUC: 0.91 vs 0.86), 3 months (AUC: 0.78 vs 0.74), and 1-year (AUC: 0.84 vs 0.79). Similarly, according to the same study, the best ML methods surpassed the TIMI risk score with greater AUC values in NSTEMI/UA patients (In-hospital: 0.89 vs 0.67, 3 months: 0.85 vs 0.67, 1-year: 0.86 vs 0.68) (Lee et al., 2021).

However, based on the result of this study, the difference in AUC was higher in NSTEMI/UA group than STEMI group when both were compared to the traditional TIMI risk score. The difference in AUC values of the NSTEMI/UA group ranged from 0.34 to 0.43 for in-hospital, 0.29 to 0.38 for 30 days and 0.35 to 0.38 for 1-year models. On the other hand, the STEMI dataset had only small difference in AUC values when compared to TIMI risk score which ranged from 0.04 to 0.08 for in-hospital, 0.02 to 0.10 for 30 days and 0.02 to 0.08 for 1-year models. The same pattern can be seen in the study by Lee et al. (2021) where the NSTEMI group had higher difference in AUC values when compared to the TIMI risk score than the STEMI group. This may be due to the fact that NSTEMI/UA has more varied clinical and pathological characteristics than STEMI, the ML models may have stronger discrimination in the NSTEMI/UA group than the traditional model (Cohen & Visveswaran, 2020; Rott & Leibowitz, 2007). STEMI is caused by a full thrombotic blockage of the infarct-related artery, whereas NSTEMI/UA is caused by incomplete coronary occlusion, coronary artery spasm, coronary embolism, myocarditis, and other conditions (Kingma, 2018). Furthermore, because of the non-parametric assumption, non-linearity, and higher-order interaction, ML-based models may outperform traditional models when evaluating complicated data. Parameter tuning was also done in the ML-based models and that may also have influenced the model performance, which may fit and perform differently in different datasets (Gibson et al., 2020).

Traditional risk stratification is mostly interested in predicting short-term mortality, with only a few suggesting 1-year mortality. Some research suggested that discrimination might be improved to predict long-term mortality after the use of ML algorithms. One recent study on 1-year mortality found that utilising Logistic Model Trees, the AUC of the prediction model could be as high as 0.901 among patients admitted to the ICU with AMI

(Barrett et al., 2019). Another study found that utilising Gradient Boosting Machine, it was possible to attain good discriminative power for 1-year mortality with an AUC of 0.898 (Sherazi et al., 2020). The results of this study found that ML models demonstrated good discrimination for 1-year mortality prediction, although the AUC value was lower than the in-hospital and 30-day mortality prediction. This could be because, unlike in other research, the 1-year mortality was defined as the mortality of those who survived at hospital release over the one-year follow-up period, rather than the cumulative mortality including in-hospital death. This is similar to the goal of this study which was to assist cardiologists in developing a treatment and management plan that took into account the risk of death after a patient was discharged.

5.2 Advantages of the best ML classifier

According to this study, SVM classifiers that used SBE technique to select variables performed better than other classifiers (RF and LR). In a research by Wallert et al. (2017), SVM outperformed RF and LR in predicting mortality after 2 years of MI for both the full model and the model with selected variables. Another study predicting in-hospital mortality for ACS patients found that SVM performed better than both LR and RF in both STEMI and NSTEMI groups at a given clinical set (Vazquez et al., 2021). This might be because SVM can handle linear and nonlinear feature space separation using kernels such as the radial basis function and avoid overfitting (Mpanya et al., 2021). Even when the training sample has some bias, SVMs can be robust, and one of the reasons for this is their capacity to offer unique solutions, where they can acquire many solutions matching to each local minima for distinct samples (Allyn et al., 2017).

LR has various drawbacks, the most significant of which is that it is based on the assumption of linearity (on a log-odds scale). Although regression models offer the benefits of being simple to construct and analyse, they are of limited use in model prediction (Frizzell et al., 2017). Since LR is a linear classifier, it performs better when dealing with linear relationships, as was the case in this study's sample, and gives probabilities that may be used to provide clinical interpretations of the results (Mansoor et al., 2017). It is also worth noting that, due to differences in the variable selection procedures, the number of variables utilised to predict mortality in LR was larger than in SVM and RF.

RF also contains a number of drawbacks that could account for this. First, each split only considers a random subset of variables (*mtry*), hence datasets with a high proportion of uninformative 'noise' variables may result in informative variables being overlooked by chance at several splits. Increasing *mtry* can help performance, but it usually comes at a high cost in terms of computation time. Second, when RFs are used to make predictions, the results are a weighted average of a portion of the data that is biassed away from extremes. This could account for some of their poor calibration (Zhang & Lu, 2012).

5.3 DL model performances evaluation

With the recent advances in DL algorithms, this study applied the DL algorithm using variables selected from the best ML algorithm (SVMvarImp-SBE-SVM) for both STEMI and NSTEMI/UA groups. On testing datasets, this study demonstrated high performance for DL models using a combination of feature selection from ML classifier algorithms. Overall, the DL models, both with and without feature selection, outperformed the ML and TIMI risk scores for the three time points in STEMI and NSTEMI/UA groups. The best models for the STEMI group were the ones with the selected variables from the best ML

model which reported the AUC value of 0.96 (with 15 variables) for in-hospital, 0.93 (with 13 variables) for 30 days, and 0.90 (with 12 variables) for 1-year mortality prediction. A similar result was demonstrated on the NSTEMI/UA group where the best models were the models which incorporated the selected variables from the ML best model with the AUC value of 0.97 (with 13 variables) for in-hospital, 0.91 (with 11 variables) for 30 days and 0.88 (with 13 variables) for 1-year mortality prediction. Even though some of the DL models' AUC with the selected variables were comparatively lower (Table 4.13) than the DL models with all variables, the models with selected variables were still chosen to be the best models because the number of variables was much smaller as compared to all the variables in the STEMI group (with 50 variables) and NSTEMI/UA group (with 39 variables). The performance of DL algorithms, in terms of AUC values, in this study was observed to be better than a similar ACS study on the Korean population using DL (STEMI = 0.91, NSTEMI = 0.87) (Kwon et al., 2019a). The authors in their study used all variables in the registry without identifying significant variables associated with ACS mortality.

DL algorithms do not require feature selection and have an unlimited number of input predictors (Kwon et al., 2019c). DL algorithms employ feature learning where it will learn all of the characteristics that have been presented and perform the tasks that have been assigned to it, such as classification and detection, in order to gain significant features that can be used to predict the outcome without reducing the number of variables. On the other hand, to increase the model predictive performance, the ML algorithm requires feature selection that reduces the number of variables in the model (Vomlel et al., 2012). The feature selection identifies variables associated with the prediction outcome. Although the DL method is a "black box", it automatically learns characteristics and can outperform traditional ML algorithms in mortality prediction (Kim et al., 2019; Kwon et al., 2019c,

2019c; Sherazi et al., 2020). Unlike ML, DL significant features are unknown, and the model is not immediately interpretable, either in terms of variable importance or the approach to deciding risk score.

5.4 TIMI risk score

The TIMI risk score was originally developed to estimate 30 days' mortality risk for STEMI patients and 14-days mortality risk for NSTEMI/UA patients (Antman et al., 2000; Lee et al., 2021; Morrow et al., 2000). In the absence of a more convenient risk score system, it has since been exploited to predict in-hospital, 30 days and 1-year mortality STEMI and NSTEMI/UA in Western and other Asian countries (Correia et al., 2014). Millo et al. (2021) conducted a study to validate the TIMI risk score for STEMI in the Caucasian population and reported the AUC value for predicting in-hospital mortality of 0.88. When the same risk score was evaluated on populations in South-East Asia countries, such as Indonesia (Martha et al., 2021) and the Philippines (Timbol et al., 2015), the AUC values for both studies reached an acceptable level of discrimination, with AUC values of 0.84 and 0.81, respectively. Another South-East Asian validation of the TIMI risk score was done by Chimparlee et al. (2018) to predict the mortality of STEMI patients for short- and long-term mortality in Thailand reported AUC value of 0.75 for in-hospital, 0.75 for 1 month, 0.77 for 6 months and 0.73 for 1 year. There is only one study that was carried out to validate the TIMI risk score for patients with STEMI in a multi-ethnic Asian population in Malaysia which reported an AUC value of 0.78 (Selvarajah et al., 2012). In this validation study, the Asian cohort was found to be carrying an overall higher disease burden and risk compared to the TIMI cohort. The mortality rate, however, was no different from this study which suggests an inherent inaccuracy within the algorithm.

The TIMI score for NSTEMI/UA has not been validated against the Western population in recent years, with the exception of a research by Bradshaw et al. (2007), which reported an AUC value of 0.80 in the prediction of 30-day mortality. Kumar et al. (2021) stated that the risk score performed well with an AUC of 0.78 when evaluated on NSTEMI patients in Pakistan. This risk score was also validated in terms of predicting long-term mortality for 6-months, 1-year, and 2-year periods, with AUC values of 0.52, 0.50, and 0.52, respectively, against Chinese population (C.-W. Chen et al., 2020). When validated against South-East Asia countries like Indonesia, a study by Karim et al. (2021) reported the performance of the TIMI risk score with an AUC value of 0.61 for NSTEMI patients and 0.63 for UA patients when predicting 14-day mortality. There is only one study that was carried out to validate the TIMI risk score for patients with NSTEMI/UA in a multi-ethnic Asian population in Malaysia which predicts 14-day mortality which reported an AUC of 0.56 (Kasim, et al., 2021a). Based on the previous studies, the performances (in terms of AUC values) of the NSTEMI/UA TIMI risk score were lower than the TIMI risk score for STEMI and this pattern was demonstrated in this study as well. This might be related to the fact that the TIMI risk score for NSTEMI/UA has not been revised since it was first introduced more than two decades ago. The performance of this conventional risk score also presented highest performance (in terms of AUC) when validated against western population. This is due to most conventional risk scores were derived from studies with mostly Caucasian patients, with just a small number of Asian patients participating in the trials. Asian patients present at a younger age with ACS, a higher prevalence of diabetes mellitus, hypertension, and renal failure, as well as a higher rate of delayed presentation for medical care are found in Asian countries compared to Western ones. Not only that, TIMI is known to underestimate mortality risk in the lower risk group.

Correia et al. (2014) reported that the TIMI score was better than the GRACE score calibration: TIMI risk score has more variables associated with death, it has a balanced distribution of patients with low, intermediate, and high-risk strata that allow specific treatment to be done on different groups of patients and the estimation done using TIMI risk score is more accurate.

However, the TIMI score has several notable limitations. First, as mentioned above, TIMI was developed using data from Western Caucasian ACS patients. Second, because TIMI risk scores only take into account the most important prognostic factors, important information may be overlooked (Kwon et al., 2019a). Exclusion of the high-risk patients is also another limitation of the risk score (Chen et al., 2018). The TIMI risk score lacks risk factors relevant to older adults and fails to account for the overall complexity of the older adult with ACS. The Asian cohort was found to be carrying an overall higher disease burden and risk compared to the TIMI cohort.

TIMI risk score for STEMI consists of the following components: age; systolic blood pressure; heart rate; Killip classification; infarct location or left bundle branch block; a history of diabetes, hypertension, angina pectoris, weight, and time to reperfusion (thrombolysis or primary PCI) while the TIMI risk score for NSTEMI/UA consists of age; equal or more than three CAD risk factors; known CAD; aspirin use in the past 7 days; severe angina; ECG ST changes more than 0.5mm; and positive cardiac marker. Previous studies have modified 'time to reperfusion' to be 'door-to-needle' or 'door-to-balloon time' instead of 'symptom onset-to-reperfusion' time because of inconsistencies in the reporting of symptom onset time (Selvarajah et al., 2013). This study excluded some variables such as angina pectoris, weight, and time to reperfusion in the model development as over 50% of data was missing as imputing a significant number of missing values might raise

problems about the outcome's acceptability and accuracy. Additional parameters (Table 1) were included including ethnicity, smoking status, invasive and non-invasive treatments, lipid profile and features from the complete blood chemistry at admission. During the development of the TIMI risk score for STEMI and NSTEMI/UA mortality prediction, continuous variables were binned. However, this study maintained the continuous variables to avoid losing potentially useful data (Shouval et al., 2017).

5.5 Selection of features using ML classifier

Feature selection algorithms are essential in mortality prediction. A combination of feature selection methods with classification algorithms resulted in higher performance versus using standalone classifiers (Aziida et al., 2021; Jafarian et al., 2011; Wallert et al., 2017). Applications of feature selection algorithms improves ML model performance using a reasonable number of predictors by reducing the predictor's dimensionality (Vomlel et al., 2012). The model performance in this study increased with the reduction in the number of predictors. This is because all of the models with selected variables had important clinical elements that contribute to predicting mortality, whereas the discarded variables did not significantly improve the model's predictive properties (Vazquez et al., 2021). According to the findings of this study, the best ML model predictive performance for STEMI patients selected 15 variables for in-hospital, 13 variables for 30 days, and 12 variables for 1-year death prediction. In the NSTEMI/UA group, 13 variables were chosen for in-hospital, 11 variables for 30 days, and 13 variables for 1-year for the optimum ML model performance. Models developed using ML and DL algorithms with the selected variables in both groups perform better than the models built using a traditional statistical method.

Since a growing number of risk factors must be monitored to estimate the risk, the system gets more complex, time-consuming, and costly as more variables are included. This increase in complexity may influence how the system is used. This study found that using variables from the feature selection method improves the prediction model's performance. This might be because all of the models with selected variables had important clinical elements that contribute to predicting mortality, whereas the discarded variables did not significantly improve the model's predictive properties in predicting the mortality of patients with ACS (Wallert et al., 2017). ML and DL models in this study were validated with untouched testing data that was not used for model development, to confirm the reliability of the current study. This study also demonstrated the ML and DL models using complete sets of variables collected, without a variable selection process resulting in an almost similar performance to models with feature selection. This shows that feature selection does not lead to the loss of important prognostic information.

Only a few research has demonstrated the impact of features on prediction model performance. One study on the 1-year mortality of patients with anterior STEMI showed a change in the performance when the top 20 ranked variables were selected instead of all 59 variables of the prediction model. The AUC for RF slightly changed from 0.932 in the full model to 0.944 with the 20 variables. When the top ranked variables were chosen, decision tree performance improved from 0.772 to 0.852, but LR performance declined from 0.931 to 0.864 (Li et al., 2020). Another study by Aziida et al. (2021) reported that their best performing models incorporated SBE and RFE as their feature selection methods in predicting 30-day mortality of ACS patients for all RF, SVM, EN, and LR algorithms. The prediction models developed using the ML algorithm appeared to be less dependent on individual predictors, owing to the fact that several clinical indicators interacted and

reflected one another. Since ML algorithms allow for non-linearity, higher-order effects, and interactions, they may not rely as much on individual predictors as traditional risk stratification methods do.

Both RF and SVM models were used to determine the list of variable importance that is an essential part of contributing to good model performance. According to this study, all the best models were a combination of SBE as the feature selection approach and SVM as the classifier (SVMvarImp-SBE-SVM). The SBE approach is frequently used in conjunction with the SVM to minimise the dimensionality of feature space (Mpanya et al., 2021). Maulidina et al. (2021) found that using SBE as a feature selection method improved the accuracy of SVM in classifying diabetes patients compared to using SVM as a classifier alone. Since the process begins with all variables included in the model, SBE provides the benefit of assessing the combined predictive capacity of variables. SBE also eliminates the least important factors from the model early on, leaving just the most essential variables (Chowdhury & Turin, 2020). RF, SVM, and LR with SBE feature reduction algorithm reported higher performance compared to RFE. SBE algorithm depends only on importance as an adequate term to eliminate unimportant variables one by one from a model. Meanwhile, RFE is reported to have poor generalisation ability (Mao, 2004).

This study also demonstrated that the final selected variables are not the top variables in the initial ranking of variable importance, but the combination of the variables throughout the ranking. It should be emphasised that the variables in the front row alone do not guarantee that the classifier will achieve the best classification results; rather, the combination of several features will ensure that the classifier will achieve the best classification results (G. Chen et al., 2020).

Based on Table 4.8 and 4.10, age, heart rate, Killip class, fasting blood glucose, and diuretics intake were ranked and selected by all short- and long-term mortality prediction ML models in the STEMI group while the common predictors for the three time points for the NSTEMI/UA group in this study were age, heart rate, Killip class and intake of Low-molecular-weight heparin (LMWH). Age, Killip class, heart rate, fasting blood glucose and diuretics intake were also selected as a factor that affects mortality post-STEMI by ML models in previous studies (Kasim et al., 2021b; Shouval et al., 2017; Wallert et al., 2017). In NSTEMI/UA patients, age, heart rate, Killip class, and LMWH were also revealed to be frequent predictors of mortality based on the previous studies (Shouval et al., 2017; Szummer et al., 2015; Weichwald et al., 2021). Older age and higher Killip class were also significant predictors of mortality (Cheng et al., 2016; Granger et al., 2003). In patients with ACS, the Killip class predicts survival, with a higher class indicating a greater risk of death (Juhan et al., 2019). Glucose levels were ranked by all selected ML models, supporting the relationship between hyperglycemia and increased risk in mortality for patients with STEMI in the Asian population (Johansson et al., 2017). STEMI and NSTEMI/UA patients with higher heart rates upon presentation are associated with an increased risk of mortality, for both short and long term (Hryhoriy, 2016; Kovar et al., 2004; Zheng et al., 2019). This may be a reflection of worse presentation (higher Killip class) or even higher pain intensity from a larger infarct. Diuretic intake is noted as one of the predictors of cardiovascular mortality in patients often given to reduce pulmonary congestion normally reflect ventricular failure and hence, a sicker group of patients (Okabe et al., 2018). For patients with a major pulmonary embolism and for the first therapy of deep vein thrombosis, LMWH is recommended and proven to lower the mortality of patients with NSTEMI/UA (He et al., 2018).

Patients' demographics such as gender and race play a big role in classifying patients' mortality for this ACS study. In a study by (Lee et al., 2013), women in Malaysia were 5 years older at presentation than males and had a higher prevalence of risk factors. Women had a greater in-hospital and six-month death rate. This is due to the fact that women are more likely to have atypical symptoms and may have waited longer to seek professional help, potentially affecting treatment options such as cardiac catheterization and PCI (Mansoor et al., 2017). Ethnicity produces a discrepancy in the gene, resulting in a unique cardiovascular risk profile marked by a high incidence of insulin resistance, glucose intolerance, central obesity, and diabetes, as well as elevated blood levels of other CAD risk markers(Chaturvedi, 2003). In Malaysia, the Malays had a higher BMI, the Chinese had a greater incidence of hypertension and hyperlipidemia, and the Indians had a higher prevalence of diabetes mellitus and a family history of early CAD (Lu & Nordin, 2013).

The high incidence of CAD risk factors such as hypertension, diabetes, dyslipidemia, smoking, and obesity is also regularly documented in the NCVD database registry in Malaysia (Ahmad, 2017). On admission for ACS, more than 95% of patients had at least one documented cardiovascular risk factor. Whereas, chronic renal disease patients have a three-fold increased risk of MI, as well as increased morbidity and death (Saad et al., 2016). On the other hand, in addition to common risk factors, chronic lung disease effects such as inflammation, endothelial dysfunction, and increased arterial stiffness are all considered to contribute to cardiovascular risk mortality (Fabbri et al., 2008; Rothnie et al., 2015). Poorly controlled CVD risk leads to an adverse systemic remodeling, leading to a plethora of cardiovascular conditions including heart failure, stroke, renal failure, and peripheral vascular disease (Rajadurai et al., 2017).

Systolic and diastolic blood pressure were ranked as predictors for both STEMI and NSTEMI/UA models. Both systolic and diastolic blood pressure help to define cardiogenic shock. Cardiogenic shock at presentation increases the risk of death. STEMI patients with cardiogenic shock who survived in-hospital death are at an increased risk of long-term death, probably as a reflection of the severity of heart attack during initial admission (Laufer-Perl et al., 2015). Cardiogenic shock is also a complication in NSTEMI patients and when compared to medical therapy alone, early revascularization is the standard treatment and is associated with improved short- and long-term survival (Kolte et al., 2016).

As lipid profiles and blood glucose levels are indications of metabolic diseases such as diabetes and hypertension, they are also critical for predicting mortality. The observed reduction in coronary events after lipid-lowering therapy is linked to changes in both low-density lipoprotein cholesterol (LDL-C) and high-density lipoprotein cholesterol (HDL-C) levels (Meeusen et al., 2017). LDL-C is an independent CV risk factor, and more Asian individuals with a very high risk of recurrent cardiovascular events had LDL-C levels above the suggested range (Poh et al., 2018; Wang & Liang, 2020). Thus, intensive lipid-lowering medication is required in ACS patients, and statins are found to be the foundation of lipid-lowering treatment in patients with ACS (Li et al., 2018; Wang & Liang, 2020). Despite the fact that lipid-lowering medication was common, it was not widely used throughout Asia (Poh et al., 2018).

Since ECG measures the electrical stability of the heart, it can be used to screen for susceptible myocardium that could lead to acute MI. The morphology of ECG complexes helps identify patients with high-risk complications after a cardiac event such as persistent ST-elevation, poor R-wave progression, and the presence of Q-waves. During the acute

event, ECG changes such as ST-elevation, ST-depression, and T-wave inversion identify patients at higher risk of death (Myers et al., 2017).

In this study, incorporating invasive or non-invasive management into the ML and DL best models for mortality prediction produced noteworthy findings. Invasive treatment, such as cardiac catheterization, is associated with improved outcomes in both STEMI and NSTEMI/UA patients (Dworeck et al., 2020; Feldkamp et al., 2018). Depending on the clinical presentation, each stage of this illness can be treated differently, however, a catheter-based interventional approach is frequently favoured. For years, antiplatelet and anticoagulant therapy, anti-anginal medicines, intensive lipid-lowering agents, and risk factor reduction are the mainstays of ACS treatment. The introduction of thrombolytics changed the way ACS was approached and treated.

Non-invasive treatment predictors such as pharmacological therapy (medications including anti-hypertensive (ACE inhibitors, beta-blockers, diuretics, ARBs), anti-diabetic agents (oral hypoglycaemic agents, insulin, and antiplatelet) were selected for in-hospital, 30 days and 1-year mortality prediction in this study. These drugs are often prescribed in the acute setting to augment neurohumoral modulation associated with left ventricular negative remodelling. Being on these medications could signal a sicker ventricle hence the strong association with death (Seong & John, 2016).

Existing risk metrics, such as the TIMI and GRACE scores, rely on lab findings that are often available within a few hours of the patient presenting with signs and symptoms of an ACS. Consequently, in this study, all the data from the best models (SVMvarImp-SBE-SVM) selected variables can be obtained in a similar time frame as the information needed to use traditional risk scores. Risk assessment within the initial few hours of presentation is

critical in managing ACS as patients identified as high risk may benefit from invasive methods within 24–48 hours after the diagnosis of ACS (Bavry et al., 2006).

5.6 Data imputation and algorithm optimisation

In this study, inadequate data led to the exclusion of nearly two-thirds of the cases. The diagnosis of ACS can be difficult, and there may be differences in how this clinical condition is reported. Some post-discharge data may not be tracked or reported directly. Data imputation was performed to ensure the validity of the findings. When data is imputed across the whole dataset, it can cause information 'leaking' between the training and testing dataset, and this is a form of principle violation (Steele et al., 2018). Hence, in this study, only the training dataset was imputed so that the same cases will not be tested against itself and cause overfitting. Multivariable imputation was employed using chained equations and PMM method for data imputation instead of using ML-based method such as missForest in this study. The data imputation method used in this study was selected as recommended in a similar study conducted on the Swedish heart registry dataset that resulted in high model performance (Wallert et al., 2017). Moreover, Solaro et al. (2018) demonstrated that the relative performance of missForest varied with the MCAR data patterns and did not show a clear advantage. Overall, the imputation accuracy and applicability of missForest are still unclear. This study initially did not include patients with more than 50% missing data as it will require data imputation, which may affect the result. It is not a limitation for the population as it is still a large dataset. As the dataset had completed dataset for all follow-up time points, generation of the risk prediction model was possible for both ML and DL. Furthermore, identifying variables associated with short- and long-term mortality prediction usage of complete cases would lead to more reliable findings as there is the possibility of biassing as a result of the imputation approach (Wallert et al.,

2017). This study went back to using an incomplete dataset and imputed data and showed almost similar results. This may be because many of the variables with low missingness were uninformative.

In terms of the clear performance convergence across models with more samples, it should be pointed out that the majority of similar ML research has utilised samples that are less than 30% of the testing dataset of those employed here. When developing models with fewer samples, they tend to be more volatile and inaccurate. As a result, the current large-scale study highlighted a potential problem of data scarcity in the creation and performance evaluation of various algorithms. This study emphasises the significance of rigorous resampling to avoid overfitting, as well as the importance of evaluating predictive models on an untouched testing set that is not subjected to model training. This study also believes that this emphasises the potential problem of data scarcity, because proper resampling and data partitioning techniques are essential for creating reliable models, and data scarcity impede both. Having said that, more data is not necessarily better (Wallert et al., 2017).

The cross-validation approach used in this study increases the efficacy of the models during model construction as it reduces the risk of model over-fitting. Also, the classification performance is highly influenced by data pre-processing and tuning of algorithms (Kesavaraj & Sukumaran, 2013). Although a model's prediction accuracy is a crucial factor to evaluate when considering whether or not to use it in a therapeutic context, the model's interpretability is also important. Readily interpretable models are more readily accepted by the medical community and, more crucially, can lead to novel ideas that can be used as the foundation for future clinical investigations (Myers et al., 2017). A pair-wise corrected resampled t-test was used to evaluate the differences between ML models'

predictive performances. The resampled t-test is a validated tool for the comparison of outcomes between two classifiers (Dietterich, 1998; Raschka, 2018).

Despite a large proportion of missing values in the original dataset, this study was still able to apply both DL and ML algorithms against TIMI risk score and compare outcomes. This is most likely due to the adoption of a hard endpoint of mortality that was unaffected by missing values. Another possibility is that the variables extracted (15 for in-hospital, 13 for 30 days and 12 for 1-year in STEMI group and 13 for in-hospital, 11 for 30 days, and 13 for 1-year in NSTEMI/UA group) was sufficient to increase the model's precision to predict death reliably.

5.7 Benefits of the research

To begin with, this study is a prospective multicentre study with a population made up of ethnic groupings such as Malays, Chinese, Indians, and other races which represent ethnicities from across Asia. Second, unlike clinical trials, this study included a diverse group of ACS patients. Finally, by establishing links with the National Registration Department for deaths, mortality occurrences were confirmed. As a result, information on mortality was recorded in the database even for individuals who were lost to follow-up.

Overall, there are two ways to interpret the significance of the reported findings. From a clinical standpoint, variables for STEMI and NSTEMI/UA mortality were assessed and ranked. Some factors can be modified and may be used as a therapeutic target. From a methodological standpoint, this study demonstrates that doctors, particularly cardiologists, can benefit from the application of ML and DL predictive algorithms in a few ways. First, it shows that a data mining approach can be used to predict outcomes in STEMI and NSTEMI/UA patients for three different time points (in-hospital, 30 days, and 1-year)

where TIMI risk score is only focused on 30-day mortality post-STEMI and 14 days mortality post-NSTEMI/UA. Second, this study shows that, in contrast to LR-based models, non-parametric algorithms may maintain good prediction performance (i.e. discrimination) even in multidimensional scenarios with dozens of variables. Third, ML methods discussed in this study were used to rank and select significant risk factors associated with short- and long-term STEMI and NSTEMI/UA mortality. Feature selection allows better interpretation of the models by restricting the scope of variables used, selecting only those clinically relevant.

With the selected variables giving the best performance in predicting mortality, and by having data continuously collected through an electronic health records system, this study will be able to allow for the adaptation of ML and DL predictive algorithms tailored to patient's risk grouping (which in this case, STEMI and NSTEMI/UA separately) eventhough the data used for the development of this algorithm are from the year 2006 to 2016 and changes in healthcare delivery may change the level of risk and outcome for patients. However, the burden of cardiovascular disease in Malaysia has been relatively stable for the past 20 years with cardiac death being the number one lead of death (Ahmad, 2022). It is safe to assume that the population risk, the type of healthcare intervention provided as well as follow up intervention remain relatively stable to allow these algorithms to be applied. The information can be subsequently integrated into hospital computer systems at the bedside for use by physicians. This tool could be implemented upon patient presentation and based on the patient's clinical history.

Finally, the findings in this study are important for government policymakers and fundholders involved in providing care for ACS patients. In comparison to developed countries, Malaysia has a dearth of hospitals with cardiac care facilities that are

appropriately equipped and resourced for primary PCIs (Ahmad, 2017). Currently, public-access cardiac care services are mostly found in large cities. The validation of the ML and DL prediction models in the Malaysian population can offer information on the distribution of high-risk ACS patients to policymakers and fundholders and can appropriately risk-stratify patients and identify those who will benefit from primary PCI, cardiac catheterization, or thrombolytics. The addition of primary PCI resources at existing hospitals may assist regions with higher numbers or proportions of high-risk patients. From the perspective of ACS management, the risk classification can be used to help the process of healthcare planning in Malaysia. As a result, the findings of this study are applicable to this country and may be relevant to countries at a comparable revolutionary stage in cardiovascular healthcare delivery.

Future studies will focus on validation of the ML and DL algorithms in real-time involving several local hospitals for continuous assessment of its reliability. Also the application of ML and DL models that are population-specific together with conventional risk scoring method that allows better outcomes in mortality prediction, communication and could increases awareness of patients that enables behavioural modifications and better management of limited resources by clinicians. The example of the web-based interface of the algorithms developed in this study is illustrated in Appendix E.

5.8 Study limitations

This study compared the performance of ML and DL best models for in-hospital, 30 days, and 1-year with a clinical prognostic model (TIMI risk score) that was designed for 14- and 30-days mortality. This study is also tailored to a multi-ethnic population and is not suitable for use in nations with a predominantly Caucasian population. Additionally, due to

the huge number of missing values in this study, some of the variables similar to the TIMI risk score were not included, and the comparison was not (or might not be) impartial. Omitting a variable could lead to a biased finding and this study attempted to reduce this effect by applying TIMI score, ML, and DL-based score to the same population. Also, this study did not compare to other risk scores such as HEART and GRACE due to the data structure and some of the variables have a large number of missing values. Finally, it is difficult to maintain control over the selection bias that exists within registries of patients. As a result, a future real-world study is required to validate the findings of this study.

CHAPTER 6: CONCLUSIONS

In conclusion, this study demonstrates the capability of ML and DL algorithms for feature selection and prediction of in-hospital, 30-day, and 1-year population-specific mortality in ACS patients. The ML and DL prediction models are suitable for this country's multi-ethnic population, and they performed better than the conventional TIMI risk score in terms of accuracy and precision, respectively. A combination of feature selection techniques and a classification algorithm allows for the reliable selection of significant variables and the improvement in model predictive performance. These algorithms could be implemented into an online system accessible for hospital use across the country, which subsequently allows for effective resource allocation and the alleviation of decision-making in the management of ACS patients. However, this study has yet to be validated against world data and should be applied in the near future.

REFERENCES

Abbas, H., Garberson, F., Glover, E., & Wall, D. P. (2017). Machine learning for early detection of autism (and other conditions) using a parental questionnaire and home video screening. 2017 IEEE International Conference on Big Data (Big Data), pp 3558-3561.

Acharya, U. R., Fujita, H., Lih, O. S., Adam, M., Tan, J. H., & Chua, C. K. (2017). Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network. *Knowledge-Based Systems*, *132*, 62-71.

Adedokun, O. A., & Burgess, W. D. (2012). Analysis of paired dichotomous data: A gentle introduction to the McNemar test in SPSS. *Journal of MultiDisciplinary Evaluation*, *8*(17), 125-131.

Ahlgren, Å. R., Hansen, F., Sonesson, B., & Länne, T. (1997). Stiffness and diameter of the common carotid artery and abdominal aorta in women. *Ultrasound in Medicine & Biology*, *23*(7), 983-988.

Ahmad, W. (2017). WA Annual Report of the NCVD-ACS Registry, 2014–2015. *National Cardiovascular Disease Database: Kuala Lumpur, Malaysia*.

Ahmad, W. (2022). WA Annual Report of the NCVD-ACS Registry, 2018–2019. *National Cardiovascular Disease Database: Kuala Lumpur, Malaysia*.

Ahmad, W. A. W., Zambahari, R., Ismail, O., Sinnadurai, J., Rosman, A., Piaw, C. S., Abidin, I. Z., & Kui-Hian, S. (2011). Malaysian national cardiovascular disease database (NCVD)–acute coronary syndrome (ACS) registry: how are we different? *CVD Prevention and Control*, *6*(3), 81-89.

Ahmed, A., & Hannan, S. A. (2012). Data mining techniques to find out heart diseases: an overview. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, *1*(4), 18-23.

Aho, K., Derryberry, D., & Peterson, T. (2014). Model selection for ecologists: the worldviews of AIC and BIC. *Ecology*, *95*(3), 631-636.

Akaike, H. (1985). "Prediction and Entropy," A celebration of statistics: the ISI Centenary Volume (eds. A.C Atkinson and S.E. Fienberg), 1-24, Springer-Verlag, New York.

Al'Aref, S. J., Singh, G., van Rosendael, A. R., Kolli, K. K., Ma, X., Maliakal, G., Pandey, M., Lee, B. C., Wang, J., & Xu, Z. (2019). Determinants of in-hospital mortality after percutaneous coronary intervention: a machine learning approach. *Journal of the American Heart Association*, *8*(5), e011160.

Allyn, J., Allou, N., Augustin, P., Philip, I., Martinet, O., Belghiti, M., Provenchere, S., Montravers, P., & Ferdynus, C. (2017). A comparison of a machine learning model with EuroSCORE II in predicting mortality after elective cardiac surgery: a decision curve analysis. *Plos One*, *12*(1), e0169772.

Almuallim, H., & Dietterich, T. G. (1991). Learning With Many Irrelevant Features. AAAI'91: Proceedings of the ninth National conference on Artificial intelligence-Volume 2 July 1991 Pages 547-552.

Amanpreet, S., Narina, T., & , & Aakanksha, S. (2016). *A review of supervised machine learning algorithms.* International Conference on Computing for Sustainable Global Development New Delhi, India. pp. 1310-1315.

Amaratunga, D., Cabrera, J., & Lee, Y.-S. (2008). Enriched random forests. *Bioinformatics*, *24*(18), 2010-2014.

Amsterdam, E. A., Wenger, N. K., Brindis, R. G., Casey Jr, D. E., Ganiats, T. G., Holmes Jr, D. R., Jaffe, A. S., Jneid, H., Kelly, R. F., & Kontos, M. C. (2014). 2014 AHA/ACC guideline for the management of patients with non–ST-elevation acute coronary syndromes: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*, *130*(25), 2354-2394.

Anand, S. S. (1999). Using ethnicity as a classification variable in health research: perpetuating the myth of biological determinism, serving socio-political agendas, or making valuable contributions to medical sciences? *Ethnicity and Health*, *4*(4), 241-244.

Andersen, H. R., Nielsen, T. T., Vesterlund, T., Grande, P., Abildgaard, U., Thayssen, P., ... & DANAMI-2 Investigators. (2003). Danish multicenter randomized study on fibrinolytic therapy versus acute coronary angioplasty in acute myocardial infarction: rationale and design of the DANish trial in Acute Myocardial Infarction-2 (DANAMI-2). *American Heart Journal*, *146*(2), 234-241.

Ang, J. C., Mirzal, A., Haron, H., & Hamed, H. N. A. (2015). Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *13*(5), 971-989.

Antman, E. M., Armstrong, P. W., Green, L. A., Halasyamani, L. K., Hochman, J. S., & Krumholz, H. M. (2008). 2007 focused update of the ACC/AHA 2004 guidelines for the management of patients with ST-elevation myocardial infarction. *Journal of the American College of Cardiology*, *51*(2), 210-247.

Antman, E. M., Cohen, M., Bernink, P. J., McCabe, C. H., Horacek, T., Papuchis, G., Mautner, B., Corbalan, R., Radley, D., & Braunwald, E. (2000). The TIMI risk score for unstable angina/non–ST elevation MI: a method for prognostication and therapeutic decision making. *Jama*, *284*(7), 835-842.

Aziida, N., Malek, S., Aziz, F., Ibrahim, K. S., & Kasim, S. (2021). Predicting 30-Day Mortality after an Acute Coronary Syndrome (ACS) using Machine Learning Methods for Feature Selection, Classification and Visualisation. *Sains Malaysiana*, *50*(3), 753-768.

Backus, B. E., Six, A. J., Kelder, J. C., Mast, T. P., van den Akker, F., Mast, E. G., Monnink, S. H., van Tooren, R. M., & Doevendans, P. A. (2010). Chest pain in the emergency room: a multicenter validation of the HEART Score. *Critical Pathways in Cardiology*, *9*(3), 164-169.

Bahi, M., & Batouche, M. (2018). Deep learning for ligand-based virtual screening in drug discovery. 2018 3rd International Conference on Pattern Analysis and Intelligent Systems (PAIS), 2018, 1-5.

Bahl, A., Hellack, B., Balas, M., Dinischiotu, A., Wiemann, M., Brinkmann, J., Luch, A., Renard, B. Y., & Haase, A. (2019). Recursive feature elimination in random forest classification supports nanomaterial grouping. *NanoImpact*, *15*, 100179.

Balakrishnan, S., Narayanaswamy, R., Savarimuthu, N., & Samikannu, R. (2008). SVM ranking with backward search for feature selection in type II diabetes databases. 2008 IEEE International Conference on Systems, Man and Cybernetics, pp. 2618-2633.

Barrett, L. A., Payrovnaziri, S. N., Bian, J., & He, Z. (2019). Building computational models to predict one-year mortality in ICU patients with acute myocardial infarction and post myocardial infarction syndrome. *AMIA Summits on Translational Science Proceedings*, *2019*, 407-416.

Barstow, C., Rice, M. S., & McDivitt, J. D. (2017). Acute coronary syndrome: diagnostic evaluation. *American Family Physician*, *95*(3), 170-177.

Bassand, J.-P., Hamm, C. W., Ardissino, D., Boersma, E., Budaj, A., Fernández-Avilés, F., Fox, K. A., Hasdai, D., Ohman, E. M., & Wallentin, L. (2007). Guidelines for the diagnosis and treatment of non-ST-segment elevation acute coronary syndromes: The Task Force for the Diagnosis and Treatment of Non-ST-Segment Elevation Acute Coronary Syndromes of the European Society of Cardiology. *European Heart Journal*, *28*(13), 1598-1660.

Bavry, A. A., Kumbhani, D. J., Rassi, A. N., Bhatt, D. L., & Askari, A. T. (2006). Benefit of early invasive therapy in acute coronary syndromes: a meta-analysis of contemporary randomized clinical trials. *Journal of the American College of Cardiology*, *48*(7), 1319-1325.

Benjamins, J. W., Hendriks, T., Knuuti, J., Juarez-Orozco, L. E., & van der Harst, P. (2019). A primer in artificial intelligence in cardiovascular medicine. *Netherlands Heart Journal*, *27*(9), 392-402.

Berrar, D. (2019). Cross-Validation. In S.Ranganathan, M. Gribskov, K.Nakai, C. Schonbach (Ed.). Encyclopedia of Bioinformatics and Computational Biology (pp. 542-545). Massachusetts, USA: Academic Press. doi: https://doi.org/10.106/B978-0-12-809633-8.20349-X.

Beswick, A., Brindle, P., Fahey, T., & Ebrahim, S. (2011). A systematic review of risk scoring methods and clinical decision aids used in the primary prevention of coronary heart disease (supplement). London, UK: Royal College of General Practitioners (UK).

Bini, S. A. (2018). Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care? *The Journal of Arthroplasty*, *33*(8), 2358-2361.

Birnbaum, Y., Chamoun, A. J., Anzuini, A., Lick, S. D., Ahmad, M., & Uretsky, B. F. (2003). Ventricular free wall rupture following acute myocardial infarction. *Coronary Artery Disease*, *14*(6), 463-470.

Blum, A. L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence*, *97*(1-2), 245-271.

Boersma, E., Pieper, K. S., Steyerberg, E. W., Wilcox, R. G., Chang, W.-C., Lee, K. L., Akkerhuis, K. M., Harrington, R. A., Deckers, J. W., & Armstrong, P. W. (2000). Predictors of outcome in patients with acute coronary syndromes without persistent ST-segment elevation: results from an international trial of 9461 patients. *Circulation*, *101*(22), 2557-2567.

Bollegala, D. (2017). Dynamic feature scaling for online learning of binary classifiers. *Knowledge-Based Systems*, *129*, 97-105.

Booth, H., & Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science*, *3*(1-2), 3-43.

Borrayo-Sanchez, G., Rosas-Peralta, M., Ramirez-Arias, E., Saturno-Chiu, G., Estrada-Gallegos, J., Parra-Michel, R., Hernandez-Garcia, H. R., Ayala-Lopez, E. A., Barraza-Felix, R., & Garcia-Rincon, A. (2018). STEMI and NSTEMI: Real-world Study in Mexico (RENASCA). *Archives of Medical Research*, *49*(8), 609-619.

Boser B.E., Guyon I.M., & V.N., V. (1992). *A training algorithm for optimal margin classifiers* Proceedings of the Fifth Annual Workshop on Computational Learning Theory, New York, NY, USA.

Bradshaw, P. J., Ko, D. T., Newman, A. M., Donovan, L. R., & Tu, J. V. (2007). Validation of the Thrombolysis In Myocardial Infarction (TIMI) risk index for predicting early mortality in a population-based cohort of STEMI and non-STEMI patients. *Canadian Journal of Cardiology*, *23*(1), 51-56.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123-140.

Breiman, L. (1999). Random Forests-Random Features. Technical Report, Statistics Dept. University of California, Berkley, USA, 567.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32.

Brown, M. L., & Kros, J. F. (2003). Data mining and the impact of missing data. *Industrial Management & Data Systems*. *103* (8), 611-621.

Brownlee, J. (2018). *Better Deep Learning: Train Faster, Reduce Overfitting, and Make Better Predictions*. Machine Learning Mastery. Melbourne, Australia.

Brownlee, J. (2020). *Data preparation for machine learning: data cleaning, feature selection, and data transforms in Python*. Melbourne, Australia: Machine Learning Mastery.

Bulluck, H., Zheng, H., Chan, M. Y., Foin, N., Foo, D. C., Lee, C. W., Lim, S. T., Sahlen, A., Tan, H. C., & Tan, J. W. (2019). Independent predictors of cardiac mortality and hospitalization for heart failure in a multi-ethnic asian ST-segment elevation myocardial infarction population treated by primary percutaneous coronary intervention. *Scientific Reports*, *9*(1), 1-14.

Bunn, A., & Korpela, M. (2020). An introduction to dplR. Seattle, WA, USA: Washington University.

Burazeri, G., Goda, A., Sulo, G., Stefa, J., Roshi, E., & Kark, J. D. (2007). Conventional risk factors and acute coronary syndrome during a period of socioeconomic transition: population-based case-control study in Tirana, Albania. *Croatian Medical Journal*, *48*(2), 225-233.

Busk, M., Maeng, M., Rasmussen, K., Kelbaek, H., Thayssen, P., Abildgaard, U., ... & Andersen, H. R. (2008). The Danish multicentre randomized study of fibrinolytic therapy vs. primary angioplasty in acute myocardial infarction (the DANAMI-2 trial): outcome after 3 years follow-up. *European Heart Journal*, *29*(10), 1259-1266.

Caelen, O. (2017). A Bayesian interpretation of the confusion matrix. *Annals of Mathematics and Artificial Intelligence*, *81*(3), 429-450.

Castro-Dominguez, Y., Dharmarajan, K., & McNamara, R. L. (2018). Predicting death after acute myocardial infarction. *Trends in Cardiovascular Medicine*, *28*(2), 102-109.

Catal, C. (2012). Performance evaluation metrics for software fault prediction studies. *Acta Polytechnica Hungarica*, *9*(4), 193-206.

Chandralekha, M., & Shenbagavadivu, N. (2018). Performance analysis of various machine learning techniques to predict cardiovascular disease: an emprical study. *Appl. Math*, *12*(1), 217-226.

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, *40*(1), 16-28.

Chattopadhyay, S., George, A., John, J., & Sathyapalan, T. (2018). Adjustment of the GRACE score by 2-hour post-load glucose improves prediction of long-term major adverse cardiac events in acute coronary syndrome in patients without known diabetes. *European Heart Journal*, *39*(29), 2740-2745.

Chaturvedi, N. (2003). Ethnic differences in cardiovascular disease. *Heart*, *89*(6), 681-686.

Chen, C.-W., Hsieh, Y.-C., Hsieh, M.-H., Lin, Y.-K., Huang, C.-Y., & Yeh, J.-S. (2020). Predictive power of in-hospital and long-term mortality of the grace, timi, revised cadillac and pami score in nstemi patients with diabetes–data from tsoc acs-dm registry. *Acta Cardiologica Sinica*, *36*(6), 595.

Chen, G., Xie, X., & Li, S. (2020). Research on Complex Classification Algorithm of Breast Cancer Chip Based on SVM-RFE Gene Feature Screening. *Complexity*, *2020,1-12*.

Chen, H., Lin, Z., Wu, H., Wang, L., Wu, T., & Tan, C. (2015). Diagnosis of colorectal cancer by near-infrared optical fiber spectroscopy and random forest. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *135*, 185-191.

Chen, Y.-H., Huang, S.-S., & Lin, S.-J. (2018). TIMI and GRACE risk scores predict both short-term and long-term outcomes in Chinese patients with acute myocardial infarction. *Acta Cardiologica Sinica*, *34*(1), 4-12.

Cheng, J. M., Helming, A. M., van Vark, L. C., Kardys, I., Den Uil, C. A., Jewbali, L. S., van Geuns, R.-J., Zijlstra, F., van Domburg, R. T., & Boersma, E. (2016). A simple risk chart for initial risk assessment of 30-day mortality in patients with cardiogenic shock from ST-elevation myocardial infarction. *European Heart Journal: Acute Cardiovascular Care*, *5*(2), 101-107.

Chimparlee, N., Chaipromprasit, J., Athisakul, S., Lertsuwunseri, V., Buddhari, W., Udayachalerm, W., & Srimahachota, S. (2018). Comparison between TIMI and GRACE Scores as predictors for short- and long-term outcome in patients with Acute ST-Elevation Myocardial Infarction. *Journal of the American College of Cardiology*, *71*(11S), A246-A246.

Cismondi, F., Fialho, A. S., Vieira, S. M., Reti, S. R., Sousa, J. M., & Finkelstein, S. N. (2013). Missing data in medical databases: Impute, delete or classify?. *Artificial Intelligence in Medicine*, *58*(1), 63-72.

Chollet, F. (2018). *Deep Learning with Python and Keras: The Practical Guide from the Developer of the Keras Library* . Wachtendonk, Germany:MITP-Verlags GmbH & Co. KG.

Chopra, A., Dimri, A., & Pradhan, T. (2017). Prediction of factors affecting amlodipine induced pedal edema and its classification. 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 1684-1689.

Chowdhury, M. Z. I., & Turin, T. C. (2020). Variable selection strategies and its importance in clinical prediction modelling. *Family Medicine and Community Health*, *8*(1).

Cohen, M., & Visveswaran, G. (2020). Defining and managing patients with non-ST-elevation myocardial infarction: Sorting through type 1 vs other types. *Clinical Cardiology*, *43*(3), 242-250.

Cooney, M. T., Dudina, A. L., & Graham, I. M. (2009). Value and limitations of existing scores for the assessment of cardiovascular risk: a review for clinicians. *Journal of the American College of Cardiology*, *54*(14), 1209-1227.

Correia, L. C., Garcia, G., Kalil, F., Ferreira, F., Carvalhal, M., Oliveira, R., Silva, A., Vasconcelos, I., Henri, C., & Noya-Rabelo, M. (2014). Prognostic value of TIMI score versus GRACE score in ST-segment elevation myocardial infarction. *Arquivos Brasileiros de Cardiologia*, *103*, 98-106.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273-297.

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods.* Cambridge, United Kingdom: Cambridge University Press.

Damman, P., van't Hof, A., Ten Berg, J., Jukema, J., Appelman, Y., Liem, A., & de Winter, R. (2017). 2015 ESC guidelines for the management of acute coronary

syndromes in patients presenting without persistent ST-segment elevation: comments from the Dutch ACS working group. *Netherlands Heart Journal*, *25*(3), 181-185.

Darabi, H. R., Tsinis, D., Zecchini, K., Whitcomb, W. F., & Liss, A. (2018). Forecasting mortality risk for patients admitted to intensive care units using machine learning. *Procedia Computer Science*, *140*, 306-313.

Das, U., Srizon, A. Y., & Hasan, M. A. M. (2020). Accurate Recognition of Coronary Artery Disease by Applying Machine Learning Classifiers. 2020 23rd International Conference on Computer and Information Technology (ICCIT), (pp. 1-6). IEEE.

Dash, M., & Liu, H. (1997). Feature selection for classification. *Intelligent data analysis*, *1*(1-4), 131-156.

de Araújo Gonçalves, P., Ferreira, J., Aguiar, C., & Seabra-Gomes, R. (2005). TIMI, PURSUIT, and GRACE risk scores: sustained prognostic value and interaction with revascularization in NSTE-ACS. *European Heart Journal*, *26*(9), 865-872.

Deo, R. C. (2015). Machine learning in medicine. *Circulation*, *132*(20), 1920-1930.

Desai, U., Martis, R. J., Nayak, C. G., Sarika, K., & Seshikala, G. (2015, December). Machine intelligent diagnosis of ECG for arrhythmia classification using DWT, ICA and SVM techniques. In *2015 Annual IEEE India Conference (INDICON)* (pp. 1-4). IEEE.

Dey, N., Ashour, A. S., & Nguyen, G. N. (2020). Recent advancement in multimedia content using deep learning. In W.B.A. Karaa, N. Dey (Ed). Mining Multimedia Documents (pp. 193-204). New York, New York, USA :Chapman and Hall.

Díaz-Uriarte, R., & De Andres, S. A. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, *7*(1), 1-13.

Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, *10*(7), 1895-1923.

Diez-Olivan, A., Del Ser, J., Galar, D., & Sierra, B. (2019). Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0. *Information Fusion*, *50*, 92-111.

DOSM. (2020). *The top 10 causes of death*. Retrieved 1 May 2021 from https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=401&bul_id=QTU5T0dKQ1g4MHYxd3ZpMzhEMzdRdz09&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT09

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (Vol. 326). New York, NY: John Wiley & Sons.

Dworeck, C., Redfors, B., Völz, S., Haraldsson, I., Angerås, O., Råmunddal, T., Ioanes, D., Myredal, A., Odenstedt, J., & Hirlekar, G. (2020). Radial artery access is associated with lower mortality in patients undergoing primary PCI: a report from the SWEDEHEART registry. *European Heart Journal: Acute Cardiovascular Care*, *9*(4), 323-332.

Efroymson, M. (1960). Multiple regression analysis. *Mathematical Methods for Digital Computers*, 191-203.

Fabbri, L., Luppi, F., Beghé, B., & Rabe, K. (2008). Complex chronic comorbidities of COPD. *European Respiratory Journal*, *31*(1), 204-212.

Faizal, A. S. M., Thevarajah, T. M., Khor, S. M., & Chang, S.-W. (2021). A review of risk prediction models in cardiovascular disease: conventional approach vs. artificial intelligent approach. *Computer Methods and Programs in Biomedicine*, 106190.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861-874.

Feder, S. L., Schulman-Green, D., Geda, M., Williams, K., Dodson, J. A., Nanna, M. G., Allore, H. G., Murphy, T. E., Tinetti, M. E., & Gill, T. M. (2015). Physicians' perceptions of the thrombolysis in myocardial infarction (TIMI) risk score in older adults with acute myocardial infarction. *Heart & Lung*, *44*(5), 376-381.

Feldkamp, T., Luedemann, M., Spehlmann, M. E., Freitag-Wolf, S., Gaensbacher, J., Schulte, K., Bajrovic, A., Hinzmann, D., Hippe, H.-J., & Kunzendorf, U. (2018). Radial access protects from contrast media induced nephropathy after cardiac catheterization procedures. *Clinical Research in Cardiology*, *107*(2), 148-157.

Fernandez-Lozano, C., Hervella, P., Mato-Abad, V., Rodríguez-Yáñez, M., Suárez-Garaboa, S., López-Dequidt, I., Estany-Gestal, A., Sobrino, T., Campos, F., & Castillo, J. (2021). Random forest-based prediction of stroke outcome. *Scientific Reports*, *11*(1), 1-12.

Fesmire, F. M., Martin, E. J., Cao, Y., & Heath, G. W. (2012). Improving risk stratification in patients with chest pain: the Erlanger HEARTS3 score. *The American Journal of Emergency Medicine*, *30*(9), 1829-1837.

Frizzell, J. D., Liang, L., Schulte, P. J., Yancy, C. W., Heidenreich, P. A., Hernandez, A. F., Bhatt, D. L., Fonarow, G. C., & Laskey, W. K. (2017). Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: comparison of machine learning and other statistical approaches. *JAMA Cardiology*, *2*(2), 204-209.

Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, *31*(14), 2225-2236.

Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation. Departmental Technical Reports (CS). 1209. Texas, USA.

Gibson, W. J., Nafee, T., Travis, R., Yee, M., Kerneis, M., Ohman, M., & Gibson, C. M. (2020). Machine learning versus traditional risk stratification methods in acute coronary syndrome: a pooled randomized clinical trial analysis. *Journal of Thrombosis and Thrombolysis*, *49*(1), 1-9.

Goff, D. C., Lloyd-Jones, D. M., Bennett, G., Coady, S., D'agostino, R. B., Gibbons, R., Greenland, P., Lackland, D. T., Levy, D., & O'donnell, C. J. (2014). 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on

Practice Guidelines. *Journal of the American College of Cardiology*, *63*(25 Part B), 2935-2959.

Goldstein, B. A., Navar, A. M., & Carter, R. E. (2017). Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. *European Heart Journal*, *38*(23), 1805-1814.

Granger, C. B., Goldberg, R. J., Dabbous, O., Pieper, K. S., Eagle, K. A., Cannon, C. P., Van de Werf, F., Avezum, A., Goodman, S. G., & Flather, M. D. (2003). Predictors of hospital mortality in the global registry of acute coronary events. *Archives of Internal Medicine*, *163*(19), 2345-2353.

Grech, E. D., & Ramsdale, D. R. (2003). Acute coronary syndrome: unstable angina and non-ST segment elevation myocardial infarction. *Bmj*, *326*(7401), 1259-1261.

Grömping, U. (2009). Variable importance assessment in regression: linear regression versus random forest. *The American Statistician*, *63*(4), 308-319.

Gurgen, F., Guler, N., & Varol, F. (2000). Antenatal fetal risk assessment by blood-flow velocity waveforms. *IEEE Engineering in Medicine and Biology Magazine*, *19*(6), 88-93.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*(1), 389-422.

Hajian-Tilaki, K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *CaspianJournal of Internal Medicine*, *4*(2), 627.

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and Techniques* (3rd ed.). Amsterdam, Netherlnds: Elsevier.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29-36.

Hasnain, M., Pasha, M. F., Ghani, I., Imran, M., Alzahrani, M. Y., & Budiarto, R. (2020). Evaluating trust prediction and confusion matrix measures for web services ranking. *IEEE Access*, *8*, 90847-90861.

He, P., Liu, Y., Wei, X., Jiang, L., Guo, W., Guo, Z., Lin, C., Tan, N., & Chen, J. (2018). Comparison of enoxaparin and unfractionated heparin in patients with non-ST-segment elevation acute coronary syndrome undergoing percutaneous coronary intervention: a systematic review and meta-analysis. *Journal of Toracic Disease*, *10*(6), 3308.

Hernandez-Suarez, D. F., Kim, Y., Villablanca, P., Gupta, T., Wiley, J., Nieves-Rodriguez, B. G., Rodriguez-Maldonado, J., Feliu Maldonado, R., da Luz Sant'Ana, I., & Sanina, C. (2019). Machine learning prediction models for in-hospital mortality after transcatheter aortic valve replacement. *Cardiovascular Interventions*, *12*(14), 1328-1338.

Hoo, F. K., Foo, Y. L., Lim, S. M. S., Ching, S. M., & Boo, Y. L. (2016). Acute coronary syndrome in young adults from a Malaysian tertiary care centre. *Pakistan Journal of Medical Sciences*, *32*(4), 841.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). New York, NY: John Wiley & Sons.

Hryhoriy, K. (2016). Peculiarities of heart rate in the patients with unstable angina and non-ST elevation myocardial infarction and concomitant diabetes mellitus type 2. *Wiadomosci Lekarskie (Warsaw, Poland: 1960)*, *69*(3 pt 2), 524-526.

Hsieh, M.-H., Lin, S.-Y., Lin, C.-L., Hsieh, M.-J., Hsu, W.-H., Ju, S.-W., Lin, C.-C., Hsu, C. Y., & Kao, C.-H. (2019). A fitting machine learning prediction model for short-term mortality following percutaneous catheterization intervention: A nationwide population-based study. *Annals of Translational Medicine*, *7*(23).

Hsieh, M. H., Hsieh, M. J., Chen, C.-M., Hsieh, C.-C., Chao, C.-M., & Lai, C.-C. (2018). Comparison of machine learning models for the prediction of mortality of patients with unplanned extubation in intensive care units. *Scientific Reports*, *8*(1), 1-7.

Huang, J., & Ling, C. X. (2005). Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on knowledge and Data Engineering*, *17*(3), 299-310.

IHME. (2020). *Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results*. Retrieved 1 May 2021 from http://ghdx.healthdata.org/record/ihme-data/gbd-2019-fatal-police-violence-by-race-and-state-1980-2019

Investigators, P. T. (1998). Inhibition of platelet glycoprotein IIb/IIIa with eptifibatide in patients with acute coronary syndromes. *New England Journal of Medicine*, *339*(7), 436-443.

IPH. (2015). National Health and Morbidity Survey 2015 (NHMS 2015). Vol. II: Non-Communicable Diseases, Risk Factors & Other Health Problems. In: Ministry of Health Malaysia Kuala Lumpur.

Ishwaran, H., & Lu, M. (2019). Standard errors and confidence intervals for variable importance in random forest regression, classification, and survival. *Statistics in Medicine*, *38*(4), 558-582.

Jadhav, A., Pramod, D., & Ramanathan, K. (2019). Comparison of performance of data imputation methods for numeric dataset. *Applied Artificial Intelligence*, *33*(10), 913-933.

Jafarian, A., Ngom, A., & Rueda, L. (2011). A novel recursive feature subset selection algorithm. 2011 IEEE 11th International Conference on Bioinformatics and Bioengineering, (pp. 78-83). IEEE.

Jain, D., & Singh, V. (2018). Feature selection and classification systems for chronic disease prediction: A review. *Egyptian Informatics Journal*, *19*(3), 179-189.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). New York, NY:Springer.

Jensen, P. B., Jensen, L. J., & Brunak, S. (2012). Mining electronic health records: towards better research applications and clinical care. *Nature Reviews Genetics*, *13*(6), 395-405.

Jerez, J. M., Molina, I., García-Laencina, P. J., Alba, E., Ribelles, N., Martín, M., & Franco, L. (2010). Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artificial Intelligence in Medicine*, *50*(2), 105-115.

Johansson, S., Rosengren, A., Young, K., & Jennings, E. (2017). Mortality and morbidity trends after the first year in survivors of acute myocardial infarction: a systematic review. *BMC Cardiovascular Disorders*, *17*(1), 1-8.

John, G. H., Kohavi, R., & Pfleger, K. (1994). Irrelevant features and the subset selection problem. In *Machine Learning Proceedings 1994* (pp. 121-129). Elsevier.

Joseph, P., Leong, D., McKee, M., Anand, S. S., Schwalm, J.-D., Teo, K., Mente, A., & Yusuf, S. (2017). Reducing the global burden of cardiovascular disease, part 1: the epidemiology and risk factors. *Circulation Research*, *121*(6), 677-694.

Jović, A., Brkić, K., & Bogunović, N. (2015). A review of feature selection methods with applications. 2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO), (pp. 1200-1205). IEEE.

Juhan, N., Khalid, Z. M., Zubairi, Y. Z., Zuhdi, A. S. M., & Ahmad, W. A. W. (2019). Risk factors of cardiovascular disease among st-elevation myocardial infarction male patients in Malaysia from 2006 to 2013. *Jurnal Teknologi*, *81*(3).

Juneja, A., Juneja, S., Kaur, S., & Kumar, V. (2021). Predicting Diabetes Mellitus With Machine Learning Techniques Using Multi-Criteria Decision Making. *International Journal of Information Retrieval Research (IJIRR)*, *11*(2), 38-52.

Kabir, M. M., Islam, M. M., & Murase, K. (2010). A new wrapper feature selection approach using neural network. *Neurocomputing*, *73*(16-18), 3273-3283.

Karatzoglou, A., Meyer, D., & Hornik, K. (2006). Support vector machines in R. *Journal of Statistical Software*, *15*(1), 1-28.

Karim, K., Fadhilah, V. K., Aulia, M. S., Pradisa, I. S., Setyasari, A., Yosephina, S., Estu, P., & Ubaidillah, N. (2021). Neutrophil to lymphocyte ratio as mortality predictor in Acute Coronary Syndrome patients. *Journal of Cardiovascular Disease Research*, *12*(3), 905-914.

Kasim, S., Malek, S., Ibrahim, K., Hiew, J., & Aziz, M. (2021a). ACS mortality prediction in Asian in-hospital patients with deep learning using machine learning feature selection. *European Heart Journal*, *42*(Supplement_1), ehab724. 3069.

Kasim, S., Malek, S., Ibrahim, K., Amir, P., & Aziz, M. (2021b). Investigating performance of deep learning and machine learning risk stratification of Asian in-hospital patients after ST-elevation myocardial infarction. *European Heart Journal-Digital Health*, *2*(4), ztab104. 3068.

Kassahun, Y., Yu, B., Tibebu, A. T., Stoyanov, D., Giannarou, S., Metzen, J. H., & Vander Poorten, E. (2016). Surgical robotics beyond enhanced dexterity instrumentation: a survey of machine learning techniques and their role in intelligent and autonomous surgical actions. *International Journal of Computer Assisted Radiology and Surgery*, *11*(4), 553-568.

Kassambara, A. (2018). *Machine learning essentials: Practical Guide in R*. France: Sthda.

Kerr, K. F., Wang, Z., Janes, H., McClelland, R. L., Psaty, B. M., & Pepe, M. S. (2014). Net reclassification indices for evaluating risk-prediction instruments: a critical review. *Epidemiology (Cambridge, Mass.)*, *25*(1), 114.

Kesavaraj, G., & Sukumaran, S. (2013). A study on classification techniques in data mining. 2013 fourth international conference on computing, communications and networking technologies (ICCCNT), (pp. 1-7). IEEE.

Khalilia, M., Chakraborty, S., & Popescu, M. (2011). Predicting disease risks from highly imbalanced data using random forest. *BMC Medical Informatics and Decision Making*, *11*(1), 1-13.

Khalill, R., Han, L., Jing, C., & Quan, H. (2009). The use of risk scores for stratification of non-ST elevation acute coronary syndrome patients. *Experimental & Clinical Cardiology*, *14*(2), e25.

Khanday, A. M. U. D., Rabani, S. T., Khan, Q. R., Rouf, N., & Din, M. M. U. (2020). Machine learning based approaches for detecting COVID-19 using clinical text data. *International Journal of Information Technology*, *12*(3), 731-739.

Khera, R., Haimovich, J., Hurley, N. C., McNamara, R., Spertus, J. A., Desai, N., ... & Krumholz, H. M. (2021). Use of machine learning models to predict death after acute myocardial infarction. *JAMA Cardiology*, *6*(6), 633-641.

Khesroh, A. A., Al-Roumi, F., Al-Zakwani, I., Attur, S., Rashed, W., & Zubaid, M. (2017). Gender differences among patients with acute coronary syndrome in the Middle East. *Heart Views: The Official Journal of the Gulf Heart Association*, *18*(3), 77.

Kim, J., & Groeneveld, P. W. (2017). Big data, health informatics, and the future of cardiovascular medicine. In (Vol. 69, pp. 899-902). Washington, DC: American College of Cardiology Foundation Washington, DC.

Kim, S. J., Cho, K. J., & Oh, S. (2017). Development of machine learning models for diagnosis of glaucoma. *Plos One*, *12*(5), e0177726.

Kim, Y. J., Saqlian, M., & Lee, J. Y. (2019). Deep learning–based prediction model of occurrences of major adverse cardiac events during 1-year follow-up after hospital discharge in patients with AMI using knowledge mining. *Personal and Ubiquitous Computing*, 1-9.

Kingma, J. (2018). Myocardial infarction: An overview of STEMI and NSTEMI physiopathology and treatment. *World Journal of Cardiovascular Diseases*, *8*(11), 498.

Kira, K., & Rendell, L. A. (1992, July). The feature selection problem: Traditional methods and a new algorithm. In *Aaai* (Vol. 2, No. 1992a, pp. 129-134).Kleinke, K. (2017). Multiple imputation under violated distributional assumptions: A systematic evaluation of the assumed robustness of predictive mean matching. *Journal of Educational and Behavioral Statistics*, *42*(4), 371-404.

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, *97*(1-2), 273-324.

Kohonen, T., Schroeder, M., & Huang, T. (2001). Maps self-organizing. Berlin, Germany: Springer Science+Business Media.

Koller, D., & Sahami, M. (1996). *Toward optimal feature selection*. California, USA: Stanford InfoLab.

Kolte, D., Khera, S., Dabhadkar, K. C., Agarwal, S., Aronow, W. S., Timmermans, R., Jain, D., Cooper, H. A., Frishman, W. H., & Menon, V. (2016). Trends in coronary angiography, revascularization, and outcomes of cardiogenic shock complicating non–ST-elevation myocardial infarction. *The American Journal of Cardiology*, *117*(1), 1-9.

Komura, D., & Ishikawa, S. (2018). Machine learning methods for histopathological image analysis. *Computational and Structural Biotechnology Journal*, *16*, 34-42.

Köse, T., Özgür, S., Coşgun, E., Keskinoğlu, A., & Keskinoğlu, P. (2020). Effect of missing data imputation on deep learning prediction performance for vesicoureteral reflux and recurrent urinary tract infection clinical study. *BioMed Research International*, *2020*.

Kovar, D., Cannon, C. P., Bentley, J. H., Charlesworth, A., & Rogers, W. J. (2004). Does initial and delayed heart rate predict mortality in patients with acute coronary syndromes? *Clinical Cardiology*, *27*(2), 80-86.

Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., & Kitai, T. (2017). Artificial intelligence in precision cardiovascular medicine. *Journal of the American College of Cardiology*, *69*(21), 2657-2664.

Kruppa, J., Ziegler, A., & König, I. R. (2012). Risk estimation and risk prediction using machine-learning methods. *Human Genetics*, *131*(10), 1639-1654.

Kuhn, M. (2009). The caret package. *Journal of Statistical Software*, *28*(5).

Kuhn, M., & Johnson, K. (2019). *Feature engineering and selection: A practical approach for predictive models*. Florida, USA:CRC Press.

Kumar, A., & Cannon, C. P. (2009, October). Acute coronary syndromes: diagnosis and management, part I. In *Mayo Clinic Proceedings* (Vol. 84, No. 10, pp. 917-938). Elsevier.

Kumar, D., Saghir, T., Zahid, M., Ashok, A., Kumar, M., Shah, A. A., Shahid, I., Ali, S., Haque, A., & Karim, M. (2021). Validity of TIMI Score for Predicting 14-Day Mortality of Non-ST Elevation Myocardial Infarction Patients. *Cureus*, *13*(1).

Kumar, R., & Indrayan, A. (2011). Receiver operating characteristic (ROC) curve for medical researchers. *Indian Pediatrics*, *48*(4), 277-287.

Kumar, V., & Minz, S. (2014). Feature selection: a literature review. *SmartCR*, *4*(3), 211-229.

Kumbhar, P., & Mali, M. (2016). A survey on feature selection techniques and classification algorithms for efficient text classification. *International Journal of Science and Research*, *5*(5), 9.

Kuo, B.-C., Ho, H.-H., Li, C.-H., Hung, C.-C., & Taur, J.-S. (2013). A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(1), 317-326.

Kwekha-Rashid, A. S., Abduljabbar, H. N., & Alhayani, B. (2021). Coronavirus disease (COVID-19) cases analysis using machine-learning applications. *Applied Nanoscience*, 1-13.

Kwon, J.-m., Jeon, K.-H., Kim, H. M., Kim, M. J., Lim, S., Kim, K.-H., Song, P. S., Park, J., Choi, R. K., & Oh, B.-H. (2019a). Deep-learning-based risk stratification for mortality of patients with acute myocardial infarction. *Plos One*, *14*(10), e0224502.

Kwon, J.-m., Kim, K.-H., Jeon, K.-H., Lee, S. E., Lee, H.-Y., Cho, H.-J., Choi, J. O., Jeon, E.-S., Kim, M.-S., & Kim, J.-J. (2019b). Artificial intelligence algorithm for predicting mortality of patients with acute heart failure. *Plos One*, *14*(7), e0219302.

Kwon, J.-m., Jeon, K.-H., Kim, H. M., Kim, M. J., Lim, S., Kim, K.-H., Song, P. S., Park, J., Choi, R. K., & Oh, B.-H. (2019c). Deep-learning-based out-of-hospital cardiac arrest prognostic system to predict clinical outcomes. *Resuscitation*, *139*, 84-91.

Lagerqvist, B., Diderholm, E., Lindahl, B., Husted, S., Kontny, F., Ståhle, E., Swahn, E., Venge, P., Siegbahn, A., & Wallentin, L. (2005). FRISC score for selection of patients for an early invasive treatment strategy in unstable coronary artery disease. *Heart*, *91*(8), 1047-1052.

Laufer-Perl, M., Shacham, Y., Letourneau-Shesaf, S., Priesler, O., Keren, G., Roth, A., & Steinvil, A. (2015). Gender-Related Mortality and In-Hospital Complications Following ST-Segment Elevation Myocardial Infarction: Data From a Primary Percutaneous Coronary Intervention Cohort. *Clinical Cardiology*, *38*(3), 145-149.

Law, M. H., Figueiredo, M. A., & Jain, A. K. (2004). Simultaneous feature selection and clustering using mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *26*(9), 1154-1166.

Le Cessie, S., & Van Houwelingen, J. C. (1992). Ridge estimators in logistic regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *41*(1), 191-201.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444.

Lee, C. Y., Hairi, N. N., Wan Ahmad, W. A., Ismail, O., Liew, H. B., Zambahari, R., Ali, R. M., Fong, A. Y. Y., Sim, K. H., & Investigators, N.-P. (2013). Are There Gender Differences in Coronary Artery Disease? The Malaysian National Cardiovascular Disease Database–Percutaneous Coronary Intervention (NCVD-PCI) Registry. *Plos One*, *8*(8), e72382.

Lee, J., Maslove, D. M., & Dubin, J. A. (2015). Personalized mortality prediction driven by electronic medical data and a patient similarity metric. *Plos One*, *10*(5), e0127428.

Lee, W., Lee, J., Woo, S.-I., Choi, S. H., Bae, J.-W., Jung, S., Jeong, M. H., & Lee, W. K. (2021). Machine learning enhances the performance of short and long-term mortality prediction model in non-ST-segment elevation myocardial infarction. *Scientific Reports*, *11*(1), 1-14.

Leong, D. P., Joseph, P. G., McKee, M., Anand, S. S., Teo, K. K., Schwalm, J.-D., & Yusuf, S. (2017). Reducing the global burden of cardiovascular disease, part 2: prevention and treatment of cardiovascular disease. *Circulation Research*, *121*(6), 695-710.

Levine, G. N., Bates, E. R., Blankenship, J. C., Bailey, S. R., Bittl, J. A., Cercek, B., Chambers, C. E., Ellis, S. G., Guyton, R. A., & Hollenberg, S. M. (2016). 2015 ACC/AHA/SCAI focused update on primary percutaneous coronary intervention for patients with ST-elevation myocardial infarction: an update of the 2011 ACCF/AHA/SCAI guideline for percutaneous coronary intervention and the 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction. *Journal of the American College of Cardiology*, *67*(10), 1235-1250.

Li, X., Liu, H., Yang, J., Xie, G., Xu, M., & Yang, Y. (2017). Using machine learning models to predict in-hospital mortality for ST-elevation myocardial infarction patients. In *MEDINFO 2017: Precision Healthcare through Informatics* (pp. 476-480). IOS Press.

Li, Y.-H., Chao, T.-H., Liu, P.-Y., Ueng, K.-C., & Yeh, H.-I. (2018). Lipid lowering therapy for acute coronary syndrome and coronary artery disease: highlights of the 2017 Taiwan lipid guidelines for high risk patients. *Acta Cardiologica Sinica*, *34*(5), 371.

Li, Y.-m., Jiang, L.-c., He, J.-j., Jia, K.-y., Peng, Y., & Chen, M. (2020). Machine learning to predict the 1-year mortality rate after acute anterior myocardial infarction in Chinese patients. *Therapeutics and Clinical Risk Management*, *16*, 1.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, *2*(3), 18-22.

Libasin, Z., Ul-Saufie, A. Z., Ahmat, H., & Shaziayani, W. N. (2020, December). Single and Multiple Imputation Method to Replace Missing Values in Air Pollution Datasets: A Review. In *IOP Conference Series: Earth and Environmental Science* (Vol. 616, No. 1, p. 012002). IOP Publishing.

Libbrecht, M. W., & Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nature Reviews Genetics*, *16*(6), 321-332.

Lin, X., Li, C., Zhang, Y., Su, B., Fan, M., & Wei, H. (2018). Selecting feature subsets based on SVM-RFE and the overlapping ratio with applications in bioinformatics. *Molecules*, *23*(1), 52.

Little, R. J., & Rubin, D. B. (2002). Bayes and multiple imputation. *Statistical Analysis with Missing Data*, 200-220.

Liu, H., & Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering*, *17*(4), 491-502.

Liul, H., Motoda, H., & Dash, M. (1998, April). A monotonic measure for optimal feature selection. In *European Conference on Machine Learning* (pp. 101-106). Berlin, Heidelberg: Springer.

Loring, Z., Mehrotra, S., Piccini, J. P., Camm, J., Carlson, D., Fonarow, G. C., Fox, K. A., Peterson, E. D., Pieper, K., & Kakkar, A. K. (2020). Machine learning does not improve upon traditional regression in predicting outcomes in atrial fibrillation: an analysis of the ORBIT-AF and GARFIELD-AF registries. *EP Europace*, *22*(11), 1635-1644.

Louppe, G., Wehenkel, L., Sutera, A., & Geurts, P. (2013). Understanding variable importances in forests of randomized trees. *Advances in Neural Information Processing Systems*, *26*, 431-439.

Lu, H. T., & Nordin, R. B. (2013). Ethnic differences in the occurrence of acute coronary syndrome: results of the Malaysian National Cardiovascular Disease (NCVD) Database Registry (March 2006-February 2010). *BMC Cardiovascular Disorders*, *13*(1), 1-14.

Lunardon, N., Menardi, G., & Torelli, N. (2014). ROSE: A Package for Binary Imbalanced Learning. *R Journal*, *6*(1).

Macias, E., Morell, A., Serrano, J., Vicario, J. L., & Ibeas, J. (2020). Mortality prediction enhancement in end-stage renal disease: A machine learning approach. *Informatics in Medicine Unlocked*, *19*, 100351.

Mahler, S. A., Miller, C. D., Hollander, J. E., Nagurney, J. T., Birkhahn, R., Singer, A. J., Shapiro, N. I., Glynn, T., Nowak, R., & Safdar, B. (2013). Identifying patients for early discharge: performance of decision rules among patients with acute chest pain. *International Journal of Cardiology*, *168*(2), 795-802.

Maldonado, S., & Weber, R. (2009). A wrapper method for feature selection using support vector machines. *Information Sciences*, *179*(13), 2208-2217.

Maley, J. H., Wanis, K. N., Young, J. G., & Celi, L. A. (2020). Mortality prediction models, causal effects, and end-of-life decision making in the intensive care unit. *BMJ Health & Care Informatics*, *27*(3).

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, *5*(9), 1315-1316.

Mansoor, H., Elgendy, I. Y., Segal, R., Bavry, A. A., & Bian, J. (2017). Risk prediction model for in-hospital mortality in women with ST-elevation myocardial infarction: a machine learning approach. *Heart & Lung*, *46*(6), 405-411.

Mao, K. Z. (2004). Orthogonal forward selection and backward elimination algorithms for feature subset selection. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *34*(1), 629-634.

Marjanović, M., Kovačević, M., Bajat, B., & Voženílek, V. (2011). Landslide susceptibility assessment using SVM machine learning algorithm. *Engineering Geology*, *123*(3), 225-234.

Maroco, J., Silva, D., Rodrigues, A., Guerreiro, M., Santana, I., & de Mendonça, A. (2011). Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC Research Notes*, *4*(1), 1-14.

Martha, J. W., Sihite, T. A., & Listina, D. (2021). The Difference in Accuracy Between Global Registry of Acute Coronary Events Score and Thrombolysis in Myocardial Infarction Score in Predicting In-Hospital Mortality of Acute ST-Elevation Myocardial Infarction Patients. *Cardiology Research*, *12*(3), 177.

Matheny, M., McPheeters, M. L., Glasser, A., Mercaldo, N., Weaver, R. B., Jerome, R. N., Walden, R., McKoy, J. N., Pritchett, J., & Tsai, C. (2011). Systematic review of cardiovascular disease risk assessment tools.

Maulidina, F., Rustam, Z., Hartini, S., Wibowo, V. V. P., Wirasati, I., & Sadewo, W. (2021, March). Feature optimization using Backward Elimination and Support Vector Machines (SVM) algorithm for diabetes classification. In *Journal of Physics: Conference Series* (Vol. 1821, No. 1, p. 012006). IOP Publishing.

Meeusen, J. W., Donato, L. J., & Jaffe, A. S. (2017). Lipid biomarkers for risk assessment in acute coronary syndromes. *Current Cardiology Reports*, *19*(6), 48.

Mehta, P., Bukov, M., Wang, C.-H., Day, A. G., Richardson, C., Fisher, C. K., & Schwab, D. J. (2019). A high-bias, low-variance introduction to machine learning for physicists. *Physics Reports*, *810*, 1-124.

Mensah, G. A., Mokdad, A. H., Ford, E. S., Greenlund, K. J., & Croft, J. B. (2005). State of disparities in cardiovascular health in the United States. *Circulation*, *111*(10), 1233-1241.

Miao, J., & Niu, L. (2016). A survey on feature selection. *Procedia Computer Science*, *91*, 919-926.

Millo, L., McKenzie, A., De la Paz, A., Zhou, C., Yeung, M., & Stouffer, G. A. (2021). Usefulness of a Novel Risk Score to Predict In-Hospital Mortality in Patients≥ 60 Years of Age with ST Elevation Myocardial Infarction. *The American Journal of Cardiology*, *154*, 1-6.

Min, S., Lee, B., & Yoon, S. (2017). Deep learning in bioinformatics. *Briefings in Bioinformatics*, *18*(5), 851-869.

Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: review, opportunities and challenges. *Briefings in Bioinformatics*, *19*(6), 1236-1246.

MOH. (2011). *Clinical Practice Guidelines: Management of Unstable Angina/Non STE Elevation Myocardial Infarction (UA/NSTEMI) 2011*. Kuala Lumpur, Malaysia: National Heart Association Malaysia.

MOH. (2019). *Clinical Practice Guideline:Management of Acute ST Segment Elevation Myocardial Infarction (STEMI) 2019.* Kuala Lumpur, Malaysia: National Heart Association Malaysia.

Mohammed, E. A., Far, B. H., & Naugler, C. (2014). Applications of the MapReduce programming framework to clinical big data analysis: current landscape and future trends. *BioData Mining*, *7*(1), 1-23.

Mohr, F. W., Morice, M.-C., Kappetein, A. P., Feldman, T. E., Ståhle, E., Colombo, A., Mack, M. J., Holmes Jr, D. R., Morel, M.-a., & Van Dyck, N. (2013). Coronary

artery bypass graft surgery versus percutaneous coronary intervention in patients with three-vessel disease and left main coronary disease: 5-year follow-up of the randomised, clinical SYNTAX trial. *The Lancet*, *381*(9867), 629-638.

Molnar, F. J., Hutton, B., & Fergusson, D. (2008). Does analysis using "last observation carried forward" introduce bias in dementia research? *CMAJ*, *179*(8), 751-753.

Morris, T. P., White, I. R., & Royston, P. (2014). Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Medical Research Methodology*, *14*(1), 1-13.

Morrow, D. A., Antman, E. M., Charlesworth, A., Cairns, R., Murphy, S. A., de Lemos, J. A., Giugliano, R. P., McCabe, C. H., & Braunwald, E. (2000). TIMI risk score for ST-elevation myocardial infarction: a convenient, bedside, clinical score for risk assessment at presentation: an intravenous nPA for treatment of infarcting myocardium early II trial substudy. *Circulation*, *102*(17), 2031-2037.

Mpanya, D., Celik, T., Klug, E., & Ntsinjana, H. (2021). Machine learning and statistical methods for predicting mortality in heart failure. *Heart Failure Reviews*, *26*(3), 545-552.

Myers, P. D., Scirica, B. M., & Stultz, C. M. (2017). Machine learning improves risk stratification after acute coronary syndrome. *Scientific Reports*, *7*(1), 1-12.

O'gara, P. T., Kushner, F. G., Ascheim, D. D., Casey Jr, D. E., Chung, M. K., De Lemos, J. A., Ettinger, S. M., Fang, J. C., Fesmire, F. M., & Franklin, B. A. (2013). 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circulation*, *127*(4), 529-555.

Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—big data, machine learning, and clinical medicine. *The New England Journal of Medicine*, *375*(13), 1216.

Okabe, T., Yakushiji, T., Kido, T., Oyama, Y., Igawa, W., Ono, M., Ebara, S., Yamashita, K., Yamamoto, M. H., & Saito, S. (2018). The association between high-dose loop diuretic use at discharge and cardiovascular mortality in patients with heart failure. *ESC Heart Failure*, *5*(1), 87-94.

Orbann, C., Sattenspiel, L., Miller, E., & Dimka, J. (2017). Defining epidemics in computer simulation models: How do definitions influence conclusions? *Epidemics*, *19*, 24-32.

Packer, M. (2017). Who should deliver medical therapy for patients with chronic heart failure? An immediate call for action to implement a community-based collaborative solution. *Circulation: Heart Failure*, *10*(8), e004332.

Panch, T., Szolovits, P., & Atun, R. (2018). Artificial intelligence, machine learning and health systems. *Journal of Global Health*, *8*(2).

Pencina, M. J., D'Agostino Sr, R. B., D'Agostino Jr, R. B., & Vasan, R. S. (2008). Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Statistics in Medicine*, *27*(2), 157-172.

Pencina, M. J., D'Agostino Sr, R. B., & Steyerberg, E. W. (2011). Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Statistics in Medicine*, *30*(1), 11-21.

Peng, S. Y., Chuang, Y. C., Kang, T. W., & Tseng, K. H. (2010). Random forest can predict 30-day mortality of spontaneous intracerebral hemorrhage with remarkable discrimination. *European Journal of Neurology*, *17*(7), 945-950.

Peng, Y., Du, X., Rogers, K. D., Wu, Y., Gao, R., & Patel, A. (2017). Predicting in-hospital mortality in patients with acute coronary syndrome in China. *The American Journal of Cardiology*, *120*(7), 1077-1083.

Perez-Riverol, Y., Kuhn, M., Vizcaíno, J. A., Hitz, M.-P., & Audain, E. (2017). Accurate and fast feature selection workflow for high-dimensional omics data. *Plos One*, *12*(12), e0189875.

Perez, A., Dennis, R. J., Gil, J. F., Rondón, M. A., & López, A. (2002). Use of the mean, hot deck and multiple imputation techniques to predict outcome in intensive care unit patients in Colombia. *Statistics in Medicine*, *21*(24), 3885-3896.

Pickering, J. W., & Endre, Z. H. (2012). New metrics for assessing diagnostic potential of candidate biomarkers. *Clinical Journal of the American Society of Nephrology*, *7*(8), 1355-1364.

Picone, I., Latessa, I., Fiorillo, A., Scala, A., Angela Trunfio, T., & Triassi, M. (2021, March). Predicting length of stay using regression and Machine Learning models in Intensive Care Unit: a pilot study. In *2021 11th international conference on biomedical engineering and technology* (pp. 52-58). Association for Computing Machinery.

Pieszko, K., Hiczkiewicz, J., Budzianowski, P., Budzianowski, J., Rzeźniczak, J., Pieszko, K., & Burchardt, P. (2019). Predicting long-term mortality after acute coronary syndrome using machine learning techniques and hematological markers. *Disease Markers*, *2019*.

Pirooznia, M., Yang, J. Y., Yang, M. Q., & Deng, Y. (2008). A comparative study of different machine learning methods on microarray gene expression data. *BMC Genomics*, *9*(1), 1-13.

Pleister, A., Selemon, H., Elton, S. M., & Elton, T. S. (2013). Circulating miRNAs: novel biomarkers of acute coronary syndrome? *Biomarkers in Medicine*, *7*(2), 287-305.

Poh, K.-K., Ambegaonkar, B., Baxter, C. A., Brudi, P., Buddhari, W., Chiang, F.-T., Horack, M., Jang, Y., Johnson, B., & Lautsch, D. (2018). Low-density lipoprotein cholesterol target attainment in patients with stable or acute coronary heart disease in the Asia-Pacific region: results from the Dyslipidemia International Study II. *European Journal of Preventive Cardiology*, *25*(18), 1950-1963.

Poldervaart, J., Langedijk, M., Backus, B., Dekker, I., Six, A., Doevendans, P., Hoes, A., & Reitsma, J. (2017). Comparison of the GRACE, HEART and TIMI score to predict major adverse cardiac events in chest pain patients at the emergency department. *International Journal of Cardiology*, *227*, 656-661.

Ponniah, J., Shamsul, A., & Adam, B. (2012). Predictors of mortality in patients with Acute Coronary Syndrome (ACS) undergoing Percutaneous Coronary Intervention (PCI): Insights from National Cardiovascular Disease Database (NCVD), Malaysia. *The Medical Journal of Malaysia*, *67*(6), 601-605.

Prajapati, G. L., & Patle, A. (2010, November). On performing classification using SVM with radial basis and polynomial kernel functions. In *2010 3rd International Conference on Emerging Trends in Engineering and Technology* (pp. 512-515). IEEE.

Rai, P., & Khanna, P. (2011). Gender classification using Wavelet Decomposition with two-way 2DPCA and SVM. *Advances in Modeling, Optimization and Computing*, 961.

Rajadurai, Zambahar R., Abdul Rahman A. R., & al., S. e. (2017). *Clinical Practices and Guidelines on Primary & Secondary Prevention of Cardiovascular Disease 2017*. Kuala Lumpur, Malaysia: CPG Secretariat.

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, *380*(14), 1347-1358.

Ralapanawa, U., Kumarasiri, P. V. R., Jayawickreme, K. P., Kumarihamy, P., Wijeratne, Y., Ekanayake, M., & Dissanayake, C. (2019). Epidemiology and risk factors of patients with types of acute coronary syndrome presenting to a tertiary care hospital in Sri Lanka. *BMC Cardiovascular Disorders*, *19*(1), 1-9.

Ramchoun, H., Idrissi, M. A. J., Ghanou, Y., & Ettaouil, M. (2016). Multilayer Perceptron: Architecture Optimization and Training. *Int. J. Interact. Multim. Artif. Intell.*, *4*(1), 26-30.

Raschka, S. (2018). Model evaluation, model selection, and algorithm selection in machine learning. *arXiv preprint arXiv:1811.12808*.

Ravì, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G.-Z. (2016). Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics*, *21*(1), 4-21.

Reddy, K., Khaliq, A., & Henning, R. J. (2015). Recent advances in the diagnosis and treatment of acute myocardial infarction. *World Journal of Cardiology*, *7*(5), 243.

Reed, G. W., Rossi, J. E., & Cannon, C. P. (2017). Acute myocardial infarction. *The Lancet*, *389*(10065), 197-210.

Rekha, G., Tyagi, A. K., & Reddy, V. K. (2019). A wide scale classification of class imbalance problem and its solutions: a systematic literature review. *Journal of Computer Science*, *15*(7), 886.929.

R Rice, L., Wong, E., & Kolter, Z. (2020, November). Overfitting in adversarially robust deep learning. In *International Conference on Machine Learning* (pp. 8093-8104). PMLR.

Ridker, P. M., Danielson, E., Fonseca, F. A., Genest, J., Gotto Jr, A. M., Kastelein, J. J., Koenig, W., Libby, P., Lorenzatti, A. J., & MacFadyen, J. G. (2008). Rosuvastatin

to prevent vascular events in men and women with elevated C-reactive protein. *New England Journal of Medicine*, *359*(21), 2195-2207.

Rojo-Álvarez, J. L., Martínez-Ramón, M., Muñoz-Marí, J., & Camps-Valls, G. (2018). *Digital signal processing with Kernel methods*. New York, NY: John Wiley & Sons.

Rothnie, K. J., Yan, R., Smeeth, L., & Quint, J. K. (2015). Risk of myocardial infarction (MI) and death following MI in people with chronic obstructive pulmonary disease (COPD): a systematic review and meta-analysis. *BMJ Open*, *5*(9), e007824.

Rott, D., & Leibowitz, D. (2007). STEMI and NSTEMI are two distinct pathophysiological entities. *European Heart Journal*, *28*(21), 2685-2685.

Royston, P., & White, I. R. (2011). Multiple imputation by chained equations (MICE): implementation in Stata. *J Stat Softw*, *45*(4), 1-20.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581-592.

Rubin, D. B. (1988, August). An overview of multiple imputation. In *Proceedings of the survey research methods section of the American statistical association* (pp. 79-84). Princeton, NJ, USA: Citeseer.

Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (1988). *Parallel distributed processing* (Vol. 1, pp. 354-362). New York: IEEE.

Saad, M., Karam, B., Faddoul, G., El Douaihy, Y., Yacoub, H., Baydoun, H., Boumitri, C., Barakat, I., Saifan, C., & El-Charabaty, E. (2016). Is kidney function affecting the management of myocardial infarction? A retrospective cohort study in patients with normal kidney function, chronic kidney disease stage III–V, and ESRD. *International Journal of Nephrology and Renovascular Disease*, *9*, 5.

Saeys, Y., Inza, I., & Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics*, *23*(19), 2507-2517.

Salomaa, V., Koukkunen, H., Ketonen, M., Immonen-R ih , P., K rj -Koskenkari, P. i., Mustonen, J., Lehto, S., Torppa, J., Lehtonen, A., & Tuomilehto, J. (2005). A new definition for myocardial infarction: what difference does it make? *European Heart Journal*, *26*(17), 1719-1725.

Sanderson, M., Bulloch, A. G., Wang, J., Williamson, T., & Patten, S. B. (2019). Predicting death by suicide using administrative health care system data: can feedforward neural network models improve upon logistic regression models? *Journal of Affective Disorders*, *257*, 741-747.

Saura, J. R. (2021). Using data sciences in digital marketing: Framework, methods, and performance metrics. *Journal of Innovation & Knowledge*, *6*(2), 92-102.

Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological Methods*, *7*(2), 147.

Schapire, R. E., & Singer, Y. (1999). Improved boosting algorithms using confidence-rated predictions. *Machine Learning*, *37*(3), 297-336.

Schlesinger, D. E., & Stultz, C. M. (2020). Deep learning for cardiovascular risk stratification. *Current Treatment Options in Cardiovascular Medicine*, *22*(8), 1-14.

Schmitt, P., Mandel, J., & Guedj, M. (2015). A comparison of six methods for missing data imputation. *Journal of Biometrics & Biostatistics*, *6*(1), 1.

Schölkopf, B., Smola, A. J., & Bach, F. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. Massachusetts, USA: MIT press.

Selvarajah, S., Fong, A. Y., Selvaraj, G., Haniff, J., Hairi, N. N., Bulgiba, A., & Bots, M. L. (2013). Impact of cardiac care variation on ST-elevation myocardial infarction outcomes in Malaysia. *The American Journal of Cardiology*, *111*(9), 1270-1276.

Selvarajah, S., Fong, A. Y. Y., Selvaraj, G., Haniff, J., Uiterwaal, C. S., & Bots, M. L. (2012). An Asian validation of the TIMI risk score for ST-segment elevation myocardial infarction. *Plos One*, *7*(7), e40249.

Seong, A. C., & John, C. K. M. (2016). A review of coronary artery disease research in Malaysia. *Med J Malaysia*, *71*(Supplement 1), 46.

Shaikhina, T., Lowe, D., Daga, S., Briggs, D., Higgins, R., & Khovanova, N. (2015). Machine learning for predictive modelling based on small data in biomedical engineering. *IFAC-PapersOnLine*, *48*(20), 469-474.

Shameer, K., Johnson, K. W., Glicksberg, B. S., Dudley, J. T., & Sengupta, P. P. (2018). Machine learning in cardiovascular medicine: are we there yet? *Heart*, *104*(14), 1156-1164.

Sherazi, S. W. A., Bae, J.-W., & Lee, J. Y. (2021). A soft voting ensemble classifier for early prediction and diagnosis of occurrences of major adverse cardiovascular events for STEMI and NSTEMI during 2-year follow-up in patients with acute coronary syndrome. *Plos One*, *16*(6), e0249338.

Sherazi, S. W. A., Jeong, Y. J., Jae, M. H., Bae, J.-W., & Lee, J. Y. (2020). A machine learning–based 1-year mortality prediction model after hospital discharge for clinical patients with acute coronary syndrome. *Health Informatics Journal*, *26*(2), 1289-1304.

Shouval, R., Hadanny, A., Shlomo, N., Iakobishvili, Z., Unger, R., Zahger, D., Alcalai, R., Atar, S., Gottlieb, S., & Matetzky, S. (2017). Machine learning for prediction of 30-day mortality after ST elevation myocardial infraction: an Acute Coronary Syndrome Israeli Survey data mining study. *International Journal of Cardiology*, *246*, 7-13.

Shuvy, M., Beeri, G., Klein, E., Cohen, T., Shlomo, N., Minha, S., & Pereg, D. (2018). Accuracy of the global registry of acute coronary events (GRACE) risk score in contemporary treatment of patients with acute coronary syndrome. *Canadian Journal of Cardiology*, *34*(12), 1613-1617.

Singh, B., Singh, H., Goel, A. D., & Singh, T. P. (2020). Study of Various Risk Factors in the Patients of Acute Coronary Syndrome. *Journal of Medical Science and Clinical reserach*, 183-189.

Singh, D., & Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, *97*, 105524.

Six, A., Backus, B., & Kelder, J. (2008). Chest pain in the emergency room: value of the HEART score. *Netherlands Heart Journal*, *16*(6), 191-196.

Six, A., Backus, B., Kingma, A., & Kaandorp, S. (2012). Consumption of diagnostic procedures and other cardiology care in chest pain patients after presentation at the emergency department. *Netherlands Heart Journal*, *20*(12), 499-504.

Solaro, N., Barbiero, A., Manzi, G., & Ferrari, P. (2018). A simulation comparison of imputation methods for quantitative data in the presence of multiple data patterns. *Journal of Statistical Computation and Simulation*, *88*(18), 3588-3619.

Somasundaram, R., & Nedunchezhian, R. (2011). Evaluation of three simple imputation methods for enhancing preprocessing of data with missing values. *International Journal of Computer Applications*, *21*(10), 14-19.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, *15*(1), 1929-1958.

Steele, A. J., Denaxas, S. C., Shah, A. D., Hemingway, H., & Luscombe, N. M. (2018). Machine learning models in electronic health records can outperform conventional survival models for predicting patient mortality in coronary artery disease. *Plos One*, *13*(8), e0202344.

Stoeldraijer, L., van Duin, C., van Wissen, L., & Janssen, F. (2013). Impact of different mortality forecasting methods and explicit assumptions on projected future life expectancy: The case of the Netherlands. *Demographic Research*, *29*, 323-354.

Sugane, H., Kataoka, Y., Otsuka, F., Nakaoku, Y., Nishimura, K., Nakano, H., Murai, K., Honda, S., Hosoda, H., & Matama, H. (2021). Cardiac outcomes in patients with acute coronary syndrome attributable to calcified nodule. *Atherosclerosis*, *318*, 70-75.

Sun, X., & Yang, Z. (2008, March). Generalized McNemar's test for homogeneity of the marginal distributions. In *SAS Global Forum*, 382, 1-10.

Szabó, G. T., Ágoston, A., Csató, G., Rácz, I., Bárány, T., Uzonyi, G., Szokol, M., Sármán, B., Jebelovszki, É., & Édes, I. F. (2021). Predictors of Hospital Mortality in Patients with Acute Coronary Syndrome Complicated by Cardiogenic Shock. *Sensors*, *21*(3), 969.

Szummer, K., Oldgren, J., Lindhagen, L., Carrero, J. J., Evans, M., Spaak, J., Edfors, R., Jacobson, S. H., Andell, P., & Wallentin, L. (2015). Association between the use of fondaparinux vs low-molecular-weight heparin and clinical outcomes in patients with non–st-segment elevation myocardial infarction. *Jama*, *313*(7), 707-716.

Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging*, *15*(1), 1-28.

Taylor, S. L., Ruhaak, L. R., Kelly, K., Weiss, R. H., & Kim, K. (2017). Effects of imputation on correlation: implications for analysis of mass spectrometry data from multiple biological matrices. *Briefings in Bioinformatics*, *18*(2), 312-320.

Thompson, B., Elish, K. O., & Steele, R. (2018, December). Machine learning-based prediction of prolonged length of stay in newborns. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 1454-1459). IEEE.

Thrane, P. G., Kristensen, S. D., Olesen, K. K., Mortensen, L. S., Bøtker, H. E., Thuesen, L., ... & Maeng, M. (2020). 16-year follow-up of the Danish Acute Myocardial Infarction 2 (DANAMI-2) trial: primary percutaneous coronary intervention vs. fibrinolysis in ST-segment elevation myocardial infarction. *European Heart Journal*, *41*(7), 847-854.

Thygesen, K. (2007). behalf of the Joint ESC/ACCF/AHA/WHF Task Force for the Redefinition of Myocardial Infarction. Universal definition of myocardial infarction. *Circulation*, *116*, 2634-2654.

Thygesen, K., Alpert, J. S., Jaffe, A. S., Chaitman, B. R., Bax, J. J., Morrow, D. A., White, H. D., & Infarction, E. G. o. b. o. t. J. E. S. o. C. A. C. o. C. A. H. A. W. H. F. T. F. f. t. U. D. o. M. (2018). Fourth universal definition of myocardial infarction (2018). *Journal of the American College of Cardiology*, *72*(18), 2231-2264.

Thygesen, K., Alpert, J. S., Jaffe, A. S., Simoons, M. L., Chaitman, B. R., & Vasileva, E. Y. (2012). Task Force for the Universal Definition of Myocardial Infarction. Third universal definition of myocardial infarction. *Nature Reviews Cardiology*, *9*(11), 620-633.

Thygesen, K., Alpert, J. S., White, H. D., & Infarction, J. E. A. A. W. T. F. f. t. R. o. M. (2007). Universal definition of myocardial infarction. *Journal of the American College of Cardiology*, *50*(22), 2173-2195.

Timbol, A. B., Villaluna, R. A., & Punzalan, F. E. (2015). 106: TIMI Risk Score for STEMI: A valid study among Filipinos for predicting in-hospital mortality. *Critical Care Medicine*, *43*(12), 28.

Torpy, J. M., Burke, A. E., & Glass, R. M. (2009). Coronary heart disease risk factors. *Jama*, *302*(21), 2388-2388.

Tran, B. X., Nghiem, S., Afoakwah, C., Ha, G. H., Doan, L. P., Nguyen, T. P., Le, T. T., Latkin, C. A., Ho, C. S., & Ho, R. C. (2020). Global mapping of interventions to improve the quality of life of patients with cardiovascular diseases during 1990–2018. *Health and Quality of Life Outcomes*, *18*(1), 1-10.

Tran, L., Chi, L., Bonti, A., Abdelrazek, M., & Chen, Y.-P. P. (2021). Mortality prediction of patients with cardiovascular disease using medical claims data under artificial intelligence architectures: validation study. *JMIR Medical Informatics*, *9*(4), e25000.

Tsai, C.-F. (2009). Feature selection in bankruptcy prediction. *Knowledge-Based Systems*, *22*(2), 120-127.

Van Buuren, S. (2011). *Multiple imputation of multilevel data*. London, UK: Routledge.

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*, 1-67.

VanHouten, J. P., Starmer, J. M., Lorenzi, N. M., Maron, D. J., & Lasko, T. A. (2014). Machine learning for risk prediction of acute coronary syndrome. In *AMIA Annual Symposium Proceedings* (Vol. 2014, p. 1940). American Medical Informatics Association.

Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, *10*(5), 988-999.

Vazquez, B., Fuentes, G., Garcia, F., Borrayo, G., & Prohias, J. (2021). Risk markers by sex and age group for in-hospital mortality in patients with STEMI or NSTEMI: an approach based on machine learning. *arXiv preprint arXiv:2101.01835*.

Verma, A. A., Murray, J., Greiner, R., Cohen, J. P., Shojania, K. G., Ghassemi, M., Straus, S. E., Pou-Prom, C., & Mamdani, M. (2021). Implementing machine learning in medicine. *CMAJ*, *193*(34), E1351-E1357.

Vernon, S. T., Coffey, S., D'Souza, M., Chow, C. K., Kilian, J., Hyun, K., Shaw, J. A., Adams, M., Roberts-Thomson, P., & Brieger, D. (2019). ST-Segment–Elevation Myocardial Infarction (STEMI) patients without standard modifiable cardiovascular risk factors—How common are they, and what are their outcomes? *Journal of the American Heart Association*, *8*(21), e013296.

Veropoulos, K., Campbell, C., & Cristianini, N. (1999, July). Controlling the sensitivity of support vector machines. In *Proceedings of the international joint conference on AI* (Vol. 55, p. 60), Penn State University.

Vink, G., Frank, L. E., Pannekoek, J., & Van Buuren, S. (2014). Predictive mean matching imputation of semicontinuous variables. *Statistica Neerlandica*, *68*(1), 61-90.

Vomlel, J., Kruzık, H., Tuma, P., Precek, J., & Hutyra, M. (2012). Machine learning methods for mortality prediction in patients with st elevation myocardial infarction. *Proceedings of WUPES*, *2012*, 204-213.

Wallert, J., Tomasoni, M., Madison, G., & Held, C. (2017). Predicting two-year survival versus non-survival after first myocardial infarction using machine learning and Swedish national register data. *BMC Medical Informatics and Decision Making*, *17*(1), 1-11.

Wang, L., Huang, Z., & Wang, R. (2021). Discrimination of cracked soybean seeds by near-infrared spectroscopy and random forest variable selection. *Infrared Physics & Technology*, *115*, 103731.

Wang, P., Cong, H., Zhang, Y., & Liu, Y. (2020). Comparison of the CAMI-NSTEMI and GRACE risk model for predicting in-hospital mortality in Chinese non-ST-segment elevation myocardial infarction patients. *Cardiology Research and Practice*, *2020*.

Wang, Q., & Liang, C. (2020). Role of Lipid-lowering therapy in low-density lipoprotein cholesterol goal attainment: Focus on patients with acute coronary syndrome. *Journal of Cardiovascular Pharmacology*, *76*(6), 658.

Weichwald, S., Candreva, A., Burkholz, R., Klingenberg, R., Räber, L., Heg, D., ... & Matter, C. M. (2021). Improving 1-year mortality prediction in ACS patients using machine learning. *European Heart Journal. Acute Cardiovascular Care*, *10*(8), 855-865.

Wessler, B. S., Lai YH, L., Kramer, W., Cangelosi, M., Raman, G., Lutz, J. S., & Kent, D. M. (2015). Clinical prediction models for cardiovascular disease: tufts predictive analytics and comparative effectiveness clinical prediction model database. *Circulation: Cardiovascular Quality and Outcomes*, *8*(4), 368-375.

WHO. (2020). *Health at a Glance: Asia/Pacific 2020 measuring progress towards universal health coverage: Measuring progress towards universal health coverage*. Paris, France: OECD Publishing.

Widera, C., Pencina, M. J., Bobadilla, M., Reimann, I., Guba-Quint, A., Marquardt, I., Bethmann, K., Korf-Klingebiel, M., Kempf, T., & Lichtinghagen, R. (2013). Incremental prognostic value of biomarkers beyond the GRACE (Global Registry of Acute Coronary Events) score and high-sensitivity cardiac troponin T in non-ST-elevation acute coronary syndrome. *Clinical Chemistry*, *59*(10), 1497-1505.

Wu, X., Gao, Y., & Jiao, D. (2019). Multi-label classification based on random forest algorithm for non-intrusive load monitoring system. *Processes*, *7*(6), 337.

Yadav, P., Joseph, D., Joshi, P., Sakhi, P., Jha, R., & Gupta, J. (2010). Clinical profile & risk factors in acute coronary syndrome. *National J Comm Med*, *1*(2), 150-151.

Zeng, J., Roussis, P. C., Mohammed, A. S., Maraveas, C., Fatemi, S. A., Armaghani, D. J., & Asteris, P. G. (2021). Prediction of Peak Particle Velocity Caused by Blasting through the Combinations of Boosted-CHAID and SVM Models with Various Kernels. *Applied Sciences*, *11*(8), 3705.

Zhang, G., & Lu, Y. (2012). Bias-corrected random forests in regression. *Journal of Applied Statistics*, *39*(1), 151-160.

Zheng, Y.-Y., Wu, T.-T., Chen, Y., Hou, X.-G., Yang, Y., Ma, X., Ma, Y.-T., Zhang, J.-Y., & Xie, X. (2019). Resting heart rate and long-term outcomes in patients with percutaneous coronary intervention: results from a 10-year follow-up of the CORFCHD-PCI study. *Cardiology Research and Practice*, *2019*.

Zhou, J., Yang, T., Liu, L., & Lu, B. (2017). Chemotherapy oxaliplatin sensitizes prostate cancer to immune checkpoint blockade therapies via stimulating tumor immunogenicity. *Molecular Medicine Reports*, *16*(3), 2868-2874.

Zou, Q., Qu, K., Luo, Y., Yin, D., Ju, Y., & Tang, H. (2018). Predicting diabetes mellitus with machine learning techniques. *Frontiers in Genetics*, *9*, 515.