PREDICTING THE ONSET OF ACUTE CORONARY SYNDROME EVENTS AND IN-HOSPITAL MORTALITY USING MACHINE LEARNING APPROACHES

SONG CHEEN

FACULTY OF SCIENCE
UNIVERSITI MALAYA
KUALA LUMPUR

2023

PREDICTING THE ONSET OF ACUTE CORONARY SYNDROME EVENTS AND IN-HOSPITAL MORTALITY USING MACHINE LEARNING APPROACHES

SONG CHEEN

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

2023

UNIVERSITI MALAYA
ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **SONG CHEEN**

Matric No: **17167295/2**

Name of Degree: **DOCTOR OF PHILOSOPHY**

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

PREDICTING THE ONSET OF ACUTE CORONARY SYNDROME EVENTS AND IN-HOSPITAL MORTALITY USING MACHINE 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: 29/12/2023

Subscribed and solemnly declared before,

Witness's Signature　　　　　　　　Date: 29/12/2023

Name:

Designation:

PREDICTING THE ONSET OF ACUTE CORONARY SYNDROME EVENTS AND IN-HOSPITAL MORTALITY USING MACHINE LEARNING APPROACHES

ABSTRACT

Acute coronary syndrome (ACS) represents a significant health concern, and its risk increases with exposure to environmental factors, particularly air pollution. Understanding this association is crucial given the increasing prevalence of air pollution in many regions, particularly in Malaysia, which is affected by air pollution. This study used a comprehensive methodology to investigate the relationship between air pollution and ACS patient outcomes utilizing machine learning (ML) algorithms, including: 1) Linear Regression, 2) Logistic Regression, 3) Support Vector Machine (SVM), 4) Random Forest (RF), 5) XGBoost, 6) Naïve Bayes (NB), and 7) Stacked Ensemble ML utilizing data from the National Cardiovascular Disease Database (NCVD) Malaysia registry and air quality data from the Department of Environment (DOE) Malaysia. The ML models for regression and classification were developed and optimized; the regression models aimed to predict ACS patients' hospitalization and mortality rates, while the classification models were designed to predict the mortality risk of ACS patients under the influence of air pollution. The regression models reported an RMSE of 1.701 (RF) for predicting hospitalization rate and 0.440 (XGBoost) for predicting cardiac mortality rate on daily basis. The classification models demonstrated an AUC of 0.843 (95% CI: 0.813 – 0.873) (RF) with the in-hospital dataset and 0.840 (95% CI: 0.828 – 0.862) (XGBoost) using the emergency dataset, outperforming the conventional TIMI risk score, and the features importance is visualized using SHAP summary plots, whereby Nitrogen Oxides (NOx) and Ozone (O_3) were identified as significant features impacting the ACS patient's outcome for hospitalization, mortality rate and mortality risk. The best-performing ML models were then integrated into the 'My Heart ACS Air' web system (https://myheartacsair.uitm.edu.my/home.php), ensuring predictions are visualized and made accessible for healthcare professionals. This web system was developed using a prototype-driven approach, emphasizing user feedback, and evaluated using the System Usability Scale (SUS). The models not only provide accurate predictions but also outperform established risk scores in the presence of air pollution. The study's findings hold relevance for Malaysia, illustrating the importance of adopting such models in regions with significant air pollution. By visualizing these predictions via a web system, healthcare professionals can gain actionable insights, potentially leading to improved patient outcomes.

Keyword: Acute coronary syndrome (ACS), Air pollution, Machine learning, Visualization, Web system, Malaysia

MERAMALKAN PERMULAAN KEJADIAN SINDROM KORONARI AKUT DAN KEMATIAN DALAM HOSPITAL MENGGUNAKAN PENDEKATAN PEMBELAJARAN MESIN

ABSTRAK

Sindrom koronari akut (ACS) merupakan satu kebimbangan kesihatan yang penting, dan risikonya meningkat dengan pendedahan kepada faktor persekitaran, terutamanya pencemaran udara. Memahami hubungan ini adalah penting memandangkan semakin meningkatnya prevalens pencemaran udara di banyak kawasan, terutamanya di Malaysia yang dipengaruhi oleh pencemaran udara. Kajian ini menggunakan metodologi yang luas untuk menyiasat hubungan antara pencemaran udara dan hasil pesakit ACS dengan menggunakan pelbagai model pembelajaran mesin: 1) Linear Regression, 2) Logistic Regression, 3) Support Vector Machine (SVM), 4) Random Forest (RF) 5) XGBoost, 6) Naïve Bayes (NB), dan 7) Stacked Ensemble Machine Learning menggunakan data dari Pangkalan Data Penyakit Kardiovaskular Kebangsaan (NCVD) Malaysia dan data kualiti udara daripada Jabatan Alam Sekitar (JAS) Malaysia. Model pembelajaran mesin untuk regresi dan pengelasan telah dibangunkan dan dioptimalkan; model regresi bertujuan untuk meramalkan kadar hospitalisasi dan kadar kematian jantung pesakit ACS, manakala model pengelasan direka untuk meramalkan risiko kematian pesakit ACS di bawah pengaruh pencemaran udara. Model regresi melaporkan RMSE sebanyak 1.701 (RF) untuk kadar hospitalisasi dan 0.440 (XGBoost) untuk kadar kematian jantung. Model pengelasan menunjukkan AUC sebanyak 0.843 (95% CI: 0.813 – 0.873) (RF) dengan set data di hospital dan 0.840 (95% CI: 0.828 – 0.862) (XGBoost) dengan set data kecemasan, melebihi konvensional risiko skor TIMI, dan kepentingan ciri-ciri divisualisasikan dengan plot ringkasan SHAP, di mana Nitrogen Oksida (NOx) dan Ozon (O_3) dikenal pasti sebagai ciri-

ciri yang signifikan mempengaruhi hasil pesakit ACS. Model ML yang berprestasi terbaik kemudian diintegrasikan ke dalam sistem web 'My Heart ACS Air' (https://myheartacsair.uitm.edu.my/home.php), memastikan ramalan divisualisasikan dan dijadikan mudah diakses untuk profesional penjagaan kesihatan. Sistem web ini dibangunkan dengan pendekatan berdasarkan prototaip, menekankan maklum balas pengguna, dan dinilai menggunakan Skala Ketergunaan Sistem (SUS). Model-model ini bukan sahaja menyediakan ramalan yang tepat tetapi juga melebihi skor risiko yang telah ditetapkan, menjadikannya alat yang bernilai untuk klinikal dan pembuat dasar. Penemuan kajian ini mempunyai kepentingan untuk Malaysia, menunjukkan kepentingan mengadopsi model seperti ini di kawasan dengan pencemaran udara yang signifikan. Dengan memvisualisasikan ramalan melalui sistem web, profesional penjagaan kesihatan boleh mendapatkan pandangan yang dapat diambil tindakan, yang berpotensi membawa kepada hasil pesakit yang ditingkatkan.

Kata kunci: Sindrom Kronari Akut (ACS), Pencemaran udara, “machine learning”, Visualisasi, Sistem web, Malaysia.

ACKOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my supervisors, Associate Professor Dr. Sorayya Malek and Professor Dr. Sazzli Kasim. I am grateful to both of my supervisors for allowing me to pursue my doctorate and for their unwavering support, guidance, and encouragement. Their knowledge has been invaluable to this research and thesis preparation. I appreciate the opportunities which have enriched my knowledge immensely. I would also like to thank Dr. Khairul, Professor Dato Dr. Wan Azman, Professor Nik Meriam, and Professor Kazuaki Negishi for their support and valuable suggestions.

The dataset provided by the National Heart Association Malaysia (NHAM) and the Department of Environment (DOE) Malaysia made this study possible. I am grateful and honored to be given the opportunity to work with national data. I also appreciate the Ministry of Science and Technology and Innovation (MOSTI) for their consistent support and belief in the relevance and importance of this project.

To the dedicated examiners and lecturers of this thesis, I extend my heartfelt thanks for their invaluable time and constructive feedback which was essential in enhancing the quality and integrity of this work.

To my family members, especially my parents, siblings, friends and loved ones, words can never describe all the sacrifices that all of you have made for me over the years. For providing me with unfailing support and continuous encouragement throughout my years of study. This accomplishment would not have been possible without them. Thank you.

Finally, I would also like to thank my peers, Dr. Cham Hui, Dr. Firdaus Aziz, Ms. Putri Nur Fatin, Ms. Nurulain Ibrahim, and Mr. Tang Jun Jie for sharing their experiences and insights throughout this academic journey.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACS	:	Acute Coronary Syndrome
AI	:	Artificial Intelligence
ANN	:	Artificial Neural Network
ANOVA	:	Analysis of Variance
antiarr	:	Anti-arrhythmic agent
API	:	Application Programming Interface
AQI	:	Air Quality Index
AUC	:	Area Under the Curve
AUROC	:	Area Under Receiving Operating Characteristics
BMI	:	Body Mass Index
BP	:	Blood Pressure
CA	:	Cardiac Arrest
CABG	:	Coronary Artery Bypass Graft
CAD	:	Coronary Artery Disease
canginapast2wk	:	Chronic angina from the past 2 weeks
CI	:	Confidence Interval
CKD	:	Chronic Kidney Disease
CM	:	Confusion Matrix
CO	:	Carbon Monoxide
COVID	:	Coronavirus Disease
CSV	:	Comma-Separated Values
CVD	:	Cardiovascular Disease
DFD	:	Data Flow Diagram

DL	:	Deep Learning
DOE	:	Department of Environment
DOSM	:	Department of Statistics Malaysia
DSDM	:	Dynamic System Development Method
DT	:	Decision Tree
ECG	:	Electrocardiogram
ecgabnormstylestelev2	:	ST segment elevation ≥2mm in ≥2 contiguous frontal leads or chest leads
ED	:	Emergency department
EHR	:	Electronic health record
EPA	:	Environment Protection Agency
FAQ	:	Frequently Asked Questions
FBG	:	Fasting Blood Glucose
FDD	:	Functional Decomposition Diagram
FN	:	False Negative
FP	:	False Positive
GIS	:	Geographic Information System
GLM	:	Generalised Linear Model
GRACE	:	Global Registry of Acute Coronary Events
HDL-C	:	High Density Lipoprotein Cholesterol
heartrate	:	Heart rate
IHD	:	Ischaemic Heart Disease
killipclass	:	Killip Class
KNN	:	K-Nearest Neighbour

LDL-C	:	Low Density Lipoprotein Cholesterol
lipidla	:	Other lipid-lowering agent
MAE	:	Mean Absolute Error
MI	:	Myocardial Infraction
ML	:	Machine Learning
MOH	:	Ministry of Health
MOSTI	:	Ministry of Science Technology and Innovation
MREC	:	Medical Research and Ethics Committee
MSE	:	Mean Square Error
NB	:	Naïve Bayes
NCVD	:	National Cardiovascular Database
NHAM	:	National Heart Association Malaysia
NHMS	:	National Health & Morbidity Survey
NMRR	:	National Medical Research Register
NOx	:	Nitrogen Oxides
NRI	:	Net Reclassification Index
NSTEMI	:	Non- ST-Elevation Myocardial Infarction
O_3	:	Ozone
OOB	:	Out Of Bag
PCA	:	Principal Component Analysis
PCI	:	Percutaneous Coronary Intervention
PM	:	Particulate Matter
ptagenotification	:	Age
RAD	:	Rapid Application Development

RCT	:	Randomized Controlled Trials
RDS	:	R Data Serialization
RF	:	Random Forest
RMSE	:	Root Mean Square Error
ROC	:	Receiver Operating Curve
ROSE	:	Random Over-Sampling Examples
SBE	:	Sequential Backward elimination
SDG	:	Sustainable Development Goals
SDLC	:	System Development Life Cycle
SDP	:	Source Data Provider
SHAP	:	SHapley Additive exPlanations
SO_2	:	Sulphur Dioxide
STEMI	:	ST-Elevation Myocardial Infarction
SUS	:	System Usability Scale
SVM	:	Support Vector Machine
TIMI	:	Thrombolysis in Myocardial Infarction
TN	:	True Negative
TP	:	True Positive
UA	:	Unstable Angina
UI	:	User Interface
WHO	:	World Health Organization
XAI	:	Explainable Artificial Intelligence
XGBoost	:	eXtreme Gradient Boosting Algorithm

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

The title of the study is 'Predicting the Onset of Acute Coronary Syndrome (ACS) Events and In-Hospital Mortality using Machine Learning Approaches', aims to examine the relationship between air pollution and ACS in Malaysia and construct predictive models applying machine learning (ML) techniques and visualize the results using geospatial map. Malaysia shares the global concern that air pollution causes to public health. High levels of air pollution have had an adverse effect on health, including an increased risk of ACS in low-middle income countries. The findings of this study contribute to the management of ACS in areas where cardiologists have become scarce, particularly in rural areas.

Chapter 1.0 provides an overview of the study, beginning with the study's background, problem statements, research questions and objectives, research scope, significance of the study, and an outline of the thesis.

1.1 Background of the Study

Air pollution is commonly defined as the presence of unwanted particulates, gases, and aerosols in the lower atmosphere (Bradstreet, 1995). According to Hertel et al. (2020), it is caused by both natural and human-induced sources and can have negative effects on human health. Certain populations, such as the elderly, children, and individuals with heart or lung conditions, may be more vulnerable to the negative impacts of air pollution. To address these negative health impacts, many countries have established regulations to reduce air pollution levels. Research suggests that reducing air pollution can lead to improved health outcomes.

According to the World Health Organization (WHO) (2022), 37% of air pollution-related premature deaths in 2019 were caused by ischemic heart disease (IHD) and stroke. The highest burden is found in Southeast Asia and Western Pacific Regions. The recent burden estimates the significant role that air pollution can cause serious cardiovascular health issues and even death. Air pollution consists of both gaseous pollutants (such as carbon monoxide, oxides of nitrogen, ozone, and sulphur dioxide) and particulate matter (PM). The presence of these pollutants has become a major concern for cardiologists and specialists in environmental medicine due to their potential negative impacts on human health.

According to previous research, ambient particulate matter (PM) in air pollution has been strongly associated with an increased risk of cardiovascular diseases (CVD) (Zhao et al., 2016; Du et al., 2016; Franchini & Mannucci, 2012; Brook et al., 2010). In Southeast Asia countries, including Malaysia, transboundary haze caused by forest fires can release acid smoke, dust, and PM into the atmosphere, contributing to public health problems. In Malaysia, outdoor air pollution is a significant contributor to the majority of deaths from heart disease (WHO, 2018). Recent studies have also demonstrated an association between nitrogen oxides (NOx) and ozone (O_3) with CVD events (Zhao et al., 2016; Chen et al., 2018; Santurtún et al., 2017).

Time lags are commonly used in air pollution studies to observe short-term and long-term exposures and reveal immediate and cumulative health effects. Short-term exposures may lead to acute respiratory issues and ACS (Samoli et al., 2008; Dockery & Pope, 1994, Gestro et al., 2020), while long-term exposures might be associated with chronic respiratory and cardiovascular disease (Zanobetti et al., 2003). This study focuses the impact of long-term and short-term exposure of air pollution on ACS hospitalization and mortality rates. Furthermore, the association between air pollution and the risk of ACS mortality was studied with the emphasis on the effects of short-term exposure (time lag 0).

Most air pollution and ACS study are based on conventional statistical methods. ML has been shown to be more effective than conventional methods in studies on CVD mortality, particularly ACS (Kasim et al., 2022a; Kasim et al., 2022b; Ke et al., 2022; Wu et al., 2021;

Aziz et al., 2021; Aziida et al., 2021; Aziz F. et al., 2019). However, limited study existed on ACS and air pollution using ML approach (Lin et al., 2021). ML are able to accurately model the complex interactions between ACS and other risk factors. As a result, this study aims to fulfil the research gap by developing a more advanced ML algorithms to model relationship between ACS and air pollution for the Southeast Asia population, particularly Malaysian.

Visualization techniques can effectively convey essential information, especially when compared to numerical values. Data visualization is a useful tool for effectively communicating and interpreting information through graphical means (Krum, 2013; Grainger et al., 2016), which is not supported by conventional statistics or ML. Google Earth provides a platform for visualizing information that can be used to disseminate information on specific sites. To the best of knowledge, there is a lack of research in the literature on the use of ML and visualization in relation to CVD and air pollution in Southeast Asia.

Hence, the aim of this study is to introduce a preliminary novel approach in integrating ML algorithms with geospatial visualization tools to analyse and present the impact of air pollution and ACS outcomes through web system utilizing the data from the National Heart Association Malaysia (NHAM) and Department of Environment (DOE), Malaysia.

The final outcome of this study is to integrate ML algorithms with visualization capability into a web system. This is an interactive web system that allows users to generate predictions, visualization, data management of ACS patients in relation with Air pollution in Malaysia. The web system can also serve as a platform for policymakers in formulating health strategies in managing ACS patients.

1.2 Overview of the Study

Ischemic heart disease (IHD) is a significant cause of hospitalization and mortality in Malaysia, A study carried out from 1985 to 2000, IHD accounted for 25% to 33% of admissions and 27% to 35% of deaths in Malaysia (Zambahari, 2004).

Recent studies have suggested that exposure to poor air quality is associated with an increased risk of developing ACS (Kuźma et al., 2021; Dominguez-Rodriguez et al., 2017; Huang et al., 2017), In Southeast Asian countries, including Malaysia, outdoor air pollution is a significant contributor to most deaths from heart disease (WHO, 2018). Thus, it is a significant public health concern that requires urgent attention.

Despite the known association between air pollution and ACS, there has been limited research conducted in Southeast Asia on this topic, especially considering that certain areas have high levels of air pollution (Rani et al., 2018; Makmom Abdullah et al., 2012). The problem is further exacerbated by transboundary haze from forest fires and industrial activities (Abdullah et al., 2020; Aghamohammadi & Isahak, 2018, Jones, 2006). In addition, there is currently no web-based system available that can predict, visualizing data and managing data related to ACS and air pollution in Malaysia.

With the emergence of ML and the development of various algorithms, such as logistic regression, support vector machine (SVM, ensemble learning (EL), etc. These techniques have become capable of capturing and analyzing complex data to produce accurate results compared to conventional statistical methods. Despite the fact that the majority of individual ML-based prediction models for mortality prediction post-ACS have outstanding performance, a number of challenging issues remain. First, no single ML algorithm is superior compared to others in the same domain. Second, the combination of multiple algorithms may provide improved performance to a single algorithm, especially in the medical field, where precise results are required, the accuracy of the outcome determines the diagnostic efficacy and patient survival rate.

Identifying risk factors for mortality improves clinical patient care. To better understand ML's "black box" nature, Shapley Additive Explanations (SHAP) were used to interpret ML model by measuring the contribution of input features to the output of a ML model at the global level.

It is important to develop a tool that can predict ACS hospitalization, ACS mortality rates, and mortality risks based on ACS and air quality features specific to the Malaysian population using ML and stacked EL. The web system developed in this study aims to fill this gap by incorporating ML and data visualization techniques to accurately predict and communicate the risk of developing ACS based on the selected features.

1.3 Problem Statements

Air pollution is the leading environmental risk factor for global health and the fourth leading cause of mortality globally (Roth et al., 2020). It is a well-established risk factor for cardiovascular morbidity and mortality (O'Toole et al., 2008), but its specific impact on ACS is still poorly understood.

In Southeast Asia, especially in Malaysia, limited research has been conducted using both conventional statistical techniques and ML methods. The burden of ACS is high in Malaysia with 20-25% of all deaths in public hospitals are attributed to coronary artery diseases with higher mortality rate reported for the 30-day mortality following myocardial infarction (Ministry of Health Malaysia, 2017). According to the WHO (2018), air pollution caused 6,251 deaths in Malaysia in 2012, and mainly attributed to heart disease.

Conventional statistical techniques have limitations in modeling the complex interactions which can be done by ML and stacked EL algorithms. Furthermore, previous studies have shown that ML and EL is more effective than traditional methods in predicting ACS mortality in Malaysia and other population specific registry (Kasim et al., 2023; Aziida et al., 2021; Aziz F. et al., 2019; Aziz F. et al., 2021; Kasim S. et al., 2022a; Kasim S. et al., 2022b; Kasim S. et al., 2021)

Due to their black-box nature, it is difficult to implement ML models in clinical medicine. Since ML models are agnostic, perturbing input and observing predictions can reveal the behavior of the underlying model (Kasim S. et al., 2022**;** Zhang et al., 2022)**.**

In addition, data visualization is crucial for effectively communicating the relationship between air pollution and heart disease. However, there is a lack of literature on the use of ML, EL and visualization in relation to CVD and air pollution in Malaysia, particularly in the context of ACS.

1.4 Research Questions

This study seeks for the answers to the following research questions:

1. What factors contribute to the occurrence of ACS in Malaysia, and what are the most significant air pollution aspects related with this condition?
2. Is it possible to develop ML and EL models that can accurately predict the hospitalization and mortality rate of ACS patients associated with ACS in Malaysia, implementing air pollution and other relevant variables?
3. Is it feasible to develop a web system with integrated ML models and data visualization techniques to better understand the relationship between air pollution and ACS in Malaysia?

1.5 Research Objectives

The study aims to develop a web system that visualizes the association between air pollution and ACS onsets in Malaysia. The system will utilize ML algorithms to predict the occurrence and outcome of ACS in the presence of air pollution features and display the results on a geographical map.

By utilizing conventional statistical methods and ML algorithms to better stratify poor outcomes in patients with ACS in the presence of air pollution. To address the research questions and achieve the aim of the study, following are the objectives:

1. To evaluate the effectiveness of ML models in assessing the effects of air pollution index with the incidences of ACS in Malaysian population.
2. To develop ML models that can predict the probability of ACS patients' mortality, the hospitalization, and mortality rate in the presence of air pollution.
3. To develop a web system that incorporates ML models and provide users with an interface to interact with the models and visualize the results on a Google map.

1.6 Scope of Research

The focus of this study is to investigate the relationship between air pollution and the onset of ACS in Malaysia, which includes its subtypes ST-elevation myocardial infarction (STEMI) and Non-ST-Elevation Myocardial Infarction/Unstable Angina (NSTEMI/UA). As for the air quality data, the variables of interest are Nitrogen Oxides (NOx), Sulfur Dioxide (SO_2), Ozone (O_3), Particulate Matter 10 (PM10). These pollutants have been identified as major contributors to air pollution and have been linked to CVD (Nogueira, 2009; Simkhovich et al., 2008; Franchini & Mannucci, 2007).

This study will employ a retrospective cohort design, utilizing data from patients diagnosed with ACS in 25 Malaysian public hospitals consider as the source data provider over the course of 12 years (2016 – April 2017) supported by the National Cardiovascular Disease Database (NCVD). The air quality data provided by the Department of Environment of Malaysia (DOE) from 2006 to April 2017 will be combined with the NCVD data to investigate the association of air pollution with ACS hospitalization rate, ACS mortality rate, and ACS patients' mortality risk.

Selected features from the NCVD attributes that will be used in developing the ML model are based on a previous published journal (Kasim S. et al., 2022) and combined with the air pollution data, to investigate its association with the ACS hospitalization rate, mortality rate and the ACS patients' mortality risk. Conventional statistical analysis was performed to examine the distribution of the data.

This study involved the development of ML models using various algorithms, such as Linear Regression, Logistic Regression, Support Vector Machine (SVM), Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Naïve Bayes (NB) and stacked ensemble learning (EL). The SHAP Explainer was used to evaluate these models to obtain insight into their prediction processes and enhance the models' transparency and accountability. The model with the greatest performance was then selected and incorporated into a web-based system for the prediction of ACS hospitalization rate, ACS mortality rate which also includes Google Maps-based visualization elements, and ACS patients' mortality calculator with the presence of air pollution.

1.7 Significant of the Study

The Sustainable Development Goals (SDGs) are a set of global goals adopted by the United Nations in 2015 to end poverty, protect the planet, and ensure peace and prosperity for all people (Robert et al., 2005). One of the SDGs is to ensure healthy lives and promote well-being for all at all ages, which relates to our study in reducing the number of deaths and illnesses from air pollution.

This study aims to investigate the relationship between air pollution and the onset of ACS in Malaysia, using visualization and ML techniques. By understanding the impact of air pollution on cardiovascular disease in Malaysia, we can contribute to the achievement of the SDG related to health and well-being. The significance of this study in relation to the SDGs is discussed below:

1. SDG 3 - Good Health and Well-being.

   By examining the impact of air pollution on ACS onset in Malaysia, this study has the potential to contribute to the goal of ensuring good health and well-being for all. If a clear relationship between air pollution and ACS onset is established, this could provide evidence for the need to reduce air pollution to protect public health and reduce the burden of ACS in Malaysia.
2. SDG 11 - Sustainable Cities and Communities.

   This study has the potential to contribute to the goal of making cities and communities inclusive, safe, resilient, and sustainable. By understanding the relationship between air pollution and ACS onset in Malaysia, policymakers and communities may be able to develop strategies for improving air quality and reducing the risk of ACS by identifying high risk areas.
3. SDG 13 - Climate Action.

   Air pollution is a major contributor to climate change, and reducing air pollution has the potential to help address this global challenge. By examining the relationship between air pollution and ACS onset in Malaysia, this study has the potential to provide evidence for the need to reduce air pollution to protect both public health and the environment.

Overall, the significant of this study in relation to the SDGs lies in its potential to contribute to the achievement of several key global goals, including improving public health, promoting sustainable and inclusive communities, and addressing climate change.

1.8 Thesis Outline

This section provides an overview of the structure of the study and summarizes the key topics covered in each chapter.

Chapter 1: Introduction. This chapter provides an overview to the research study studying on the impact of air pollution on the onset of ACS in Malaysia. This chapter begins with the background of the study, followed by the overview of the study, research objectives, and problem statements. The scope of the study and its significance are then discussed, highlighting the relevance and importance of this research. The chapter concludes by presenting a thesis outline that summarizes the structure of the study and its key components.

Chapter 2: Literature Review. This chapter discusses in detail of all the relevant topic to this study, including ACS, Air Pollution, the effect of air pollution towards onset of ACS in Malaysia. Furthermore, it provides an extensive review of existing literature studies on ACS hospitalization, ACS mortality rates and ACS mortality risk, ML techniques, and web-based implementation of predictive models.

Chapter 3: Research Methodology. This chapter elaborates on the research data, summary statistics, and methodology used to design and develop the ML and stacked EL models for predicting ACS's hospitalization and mortality rates, as well as mortality risks in ACS patients with air pollution features. Additionally, the chapter covers the development of the web system prototype and data visualization using geospatial maps, including the design of the user interface, hardware and software requirements, and system usability testing.

Chapter 4: Result. This chapter outlines the findings of the regression and classification models, including graphical visualizations, comparisons with TIMI risk score, Net Reclassification Index (NRI), SHAP analysis. Furthermore, the development and functionalities of the web-based prototype – My Heart ACS Air are highlighted.

Chapter 5: Discussions. In this chapter presents analysis of the results, as well as an interpretation of the outcomes. This comprises conducting additional analysis of the model, evaluating the performance of the best ML model, and providing the findings of the web system prototype validation and system usability testing. The significance of the study and limitations are also included in this chapter.

Chapter 6: Conclusion. This chapter summarized the findings of the entire study, including its strengths, limitations, and future improvements.

CHAPTER 2: LITERATURE REVIEW

This literature review addresses previous research on air pollution and acute coronary syndrome (ACS), as well as machine learning (ML) techniques for predicting cardiovascular outcomes. We will also investigate research on health data visualization using maps and web applications.

2.1 Acute Coronary Syndrome (ACS)

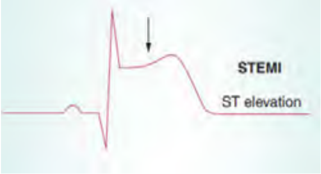
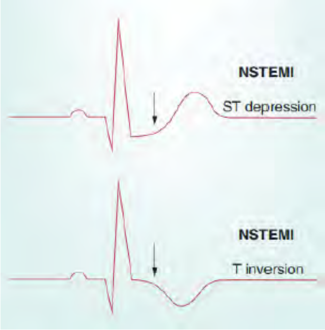
Acute Coronary Syndrome (ACS) is a type of cardiovascular disease (CVD) that involves a sudden reduction or blockage of blood flow to the heart due to narrowed or blocked coronary arteries (Ghaffari, 2022; Overbaugh, 2009). ACS often triggered by acute changes, including superficial erosion or rupture of coronary atherosclerotic plaques, which caused 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 (Amsterdam, et al., 2014; Buja & Butany, 2022).

ACS encompasses a spectrum of clinical manifestations, including unstable angina (UA) and myocardial infarction (MI), which are further divided into ST-segment elevation myocardial infarction (STEMI) and non-ST-segment elevation myocardial infarction (NSTEMI).

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). Chest pain of STEMI begins abruptly and lasts for more than thirty minutes. It is usually located in the center of the chest and may radiate to the jaw or down the left arm. It may occur at rest or with activity (Reigle, 2005).

If an electrocardiogram (ECG) does not show ST-elevation, patients may be diagnosed with either NSTEMI or UA. NSTEMI is characterized by persistent symptoms of cardiac ischemia and elevated cardiac markers, while UA presents with similar symptoms but without an elevation in cardiac troponin levels. The chest pain experienced with NSTEMI/UA is typically located in the center or left side of the chest and may radiate to the jaw or upper limb (Daga, et al., 2011). Table 2.1 summarized types of ACS and in Figure 2.1, provides an overview of ACS and its classification.

Table 2.1: Types of ACS

Types of ACS	Definition	Symptoms	Diagnosis		Illustration
			ST-segment elevation in ECG	Bio-maker	
ST-Elevation Myocardial Infarction (STEMI)	Complete occlusion of a coronary artery causing ischemia and myocardial infarction	Severe and persistent chest pain and discomfort	Yes	Yes	 (Pleister, et al., 2013)
Non-ST Elevation Myocardial Infarction (NSTEMI)	Partial occlusion of a coronary artery causing ischemia and myocardial damage	Severe and prolonged chest pain and discomfort	Yes	No	 (Pleister, et al., 2013)
Unstable Angina (UA)	New onset chest pain or change in pattern of previously stable angina	Chest pain and discomfort	No	No	N/A

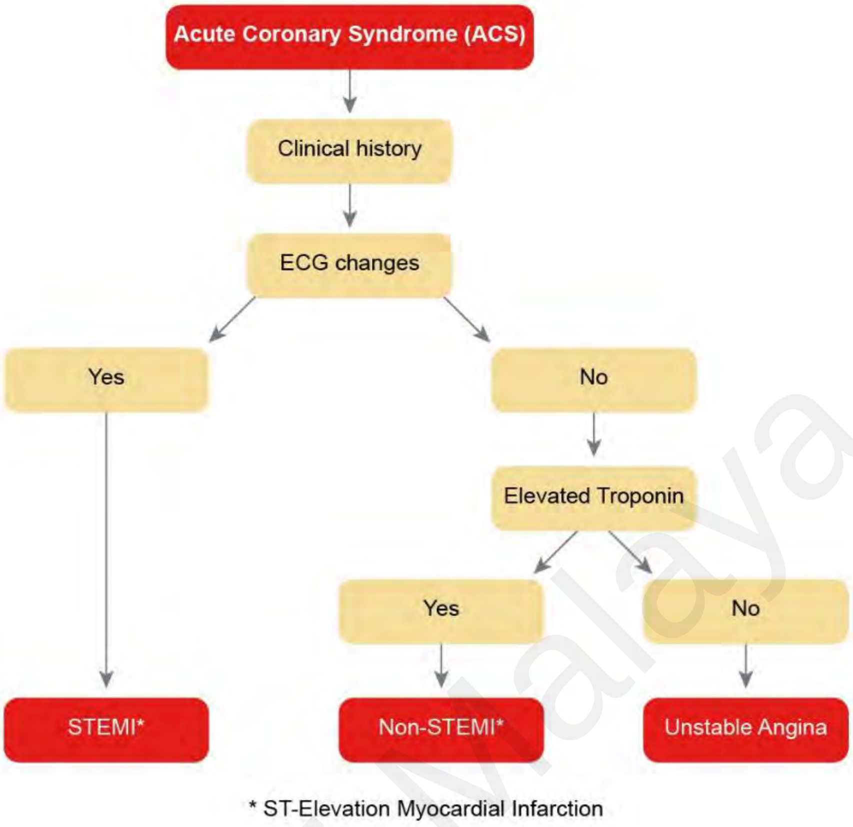


Figure 2.1: ACS and its Classification (Photo sourced from Chew, et al., 2016).

ACS is commonly known as ischemic heart disease (IHD). The World Health Organization (WHO) (2022) stated that CVD (such as IHD and stroke) is accountable for most noncommunicable diseases, which is 17.9 million annually (Roth, 2018). Furthermore, it is projected that cardiovascular disease mortality will increase from 17 million in 2008 to 25 million in 2030 annually, making cardiovascular death as the major contributor to global morbidity and mortality (Karageorgou, et al., 2015). According to the most recent statistics reported by the Department of Statistics Malaysia (DOSM) (2022), ischemic heart disease, continued to be the leading cause of death in noncommunicable diseases accounting for 13.7% in our country in 2021. According to the National Heart Association Malaysia (NHAM)'s Annual Report of the Acute Coronary Syndrome (ACS) Registry, 2018–2019, 20,605 patients are admitted with ACS (Wan Ahmad., 2022).

The burden of ACS is high in Malaysia and other similar middle-income countries. In Malaysia, 17.2% of all deaths in public hospitals are attributed to ACS and is the primary cause of death in Malaysia (DOSM, 2021). Outdoor air pollution caused 6,251 deaths in Malaysia in 2012, according to report by the WHO (2018), where the cause of death was due to heart disease (3,630), stroke (1773), lung cancer (670), pulmonary disease (148) and lower respiratory disease (29). Figure 2.2 depicts the IHD remains the primary causes of death in Malaysia in year 2020 adopted from Department of Statistics Malaysia (DOSM, 2021). In Figure 2.3 illustrates horizontal bar chart of medically certified causes of death in Malaysia, 2020-2021.

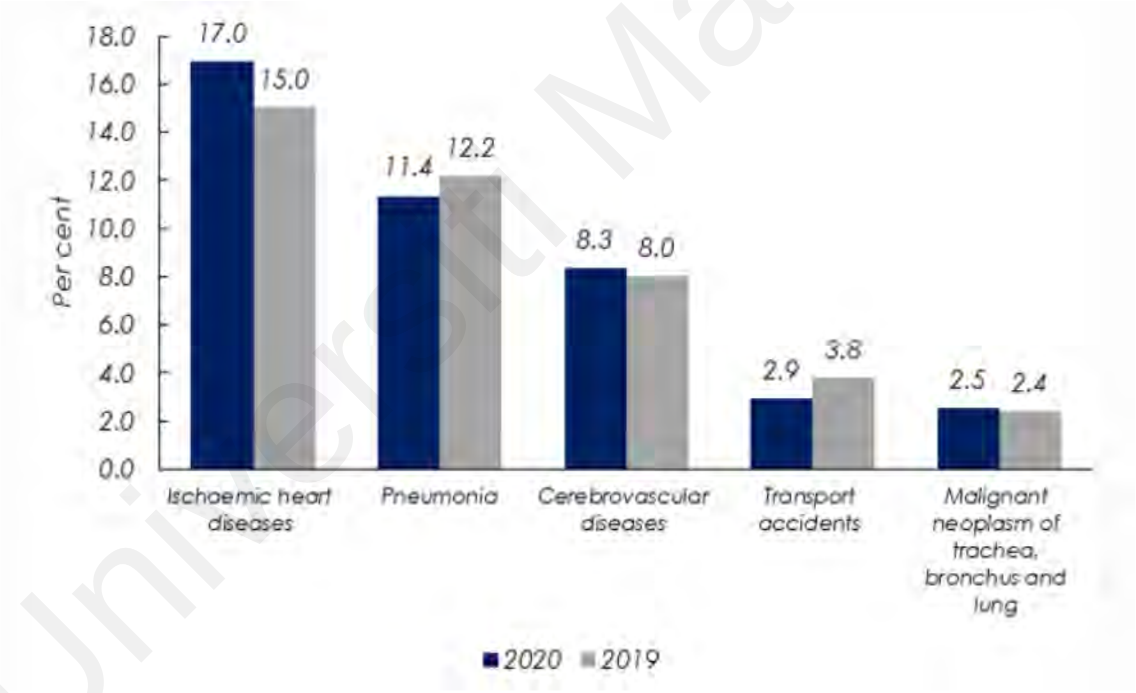


Figure 2.2: Ischemic heart disease remains the primary causes of death in Malaysia in year 2020 (Photo sourced from Department of Statistics Malaysia (DOSM), 2021).

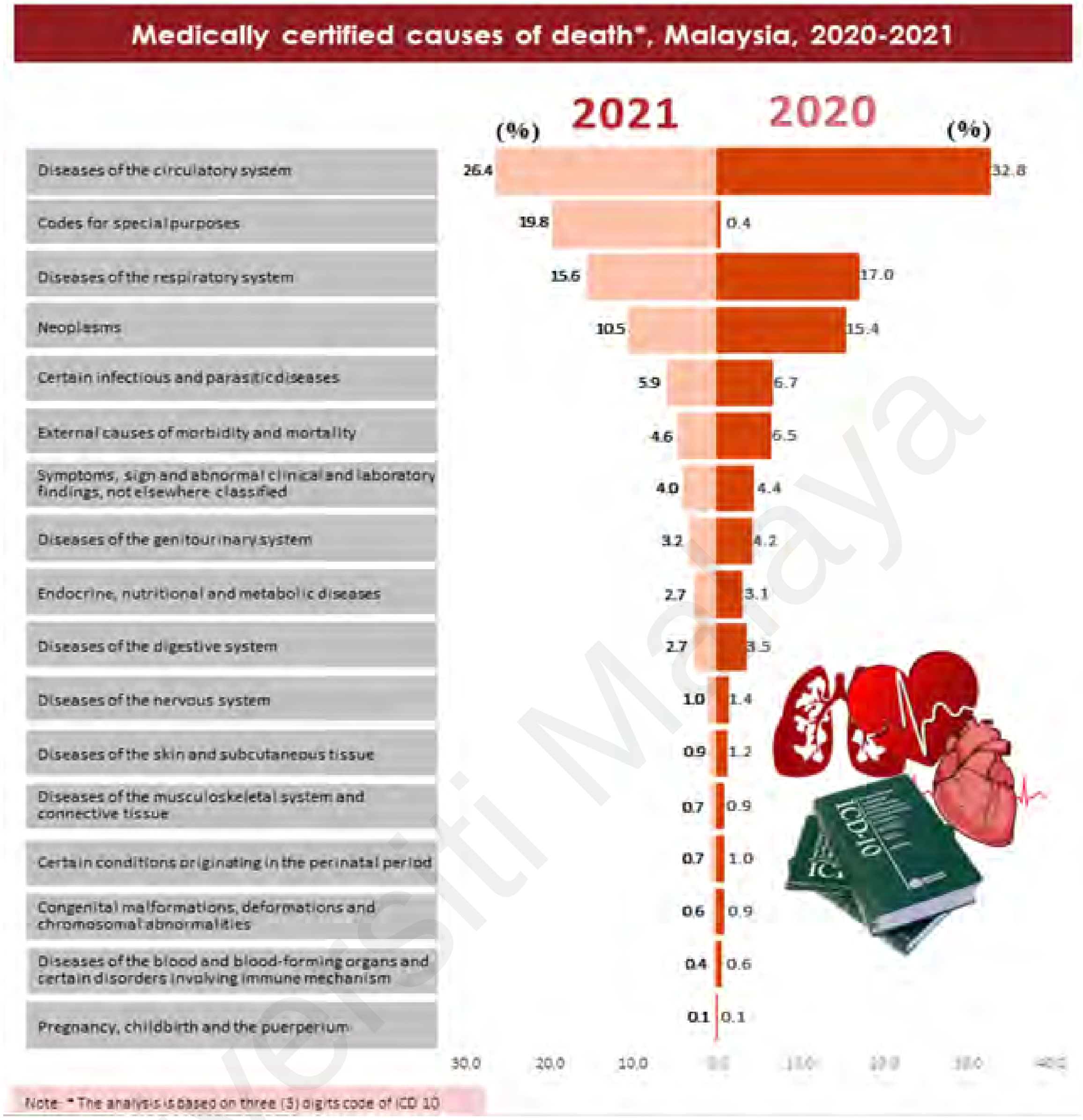


Figure 2.3: Diseases of the circulatory system ranked first in medically certified causes of death in Malaysia, 2020-2021 (Photo source from Department of Statistics Malaysia (DOSM) 2022),

2.2 Risk Factors

ACS is a multifactorial disease that results from complex interactions between genetic and environmental factors (Talmor-Barkan, et al., 2022; Roth, et al., 2020; Nansseu, et al., 2015). Risk factors stratification and identification is crucial for devising effective prevention and treatment strategies to reduce the incidence and burden of ACS by collecting information from patients about pain characteristics and symptoms, risk factors or a history of CVD, and recent medications (Goswami, et al., 2012).

Study carried out by Ralapanawa, et al. (2019), stated that despite major advances in management, IHD remains the most common kind of heart disease and the leading cause of early mortality worldwide. Several studies on epidemiology, risk factors, and outcomes of ACS in Western countries have been published. Studies indicate that South Asians have higher mortality rates and premature deaths attributable to IHD (deaths occurring at least 10–15 years earlier than expected) than individuals in Western nations (Hughes, et al., 1989; Ghaffar, et al., 2004). Unfortunately, there are only a limited number of studies that investigate the risk factors of ACS in the Southeast Asia population. Therefore, it is crucial to identify the risk factors for the Southeast Asia population, especially in Malaysia where the impact of air pollution on ACS is not yet fully studied.

Risk factors can be classified as either modifiable or non-modifiable. In general context, modifiable risk factors include smoking, hypertension, diabetes, dyslipidemia, physical inactivity, and obesity (Arnett, et al., 2019). Non-modifiable risk factors are irreversible, include age, gender, and family history of cardiovascular disease (Khera & Kathiresan, 2017).

Several modifiable risk factors for ACS have been identified, such as hypertension, smoking, dyslipidemia, diabetes, obesity, physical inactivity, and poor dietary practices (Kong, et al., 2023; Cheema, et al., 2020; Hadjiev, et al., 2003). Hypertension has been identified as a major risk factor for heart disease and stroke in Asia, where stroke morbidity and mortality rates are exceedingly high compared to Western countries (Chen, et al., 2018). In addition, Asian men have a high prevalence of smoking and diabetes, whereas cholesterol levels in Asian countries are generally lower than in Western nations (Chen, et al., 2018;

Fuster & Kelly, 2010). Obesity is also significant risk factor for cardiovascular disease in the Asian population, as determined by research conducted in Malaysia and Singapore (Zheng, et al., 2020).

CVD burden attributable to modifiable risk factors continues to increase globally. Air pollution is the leading environmental risk factor for global health and the fourth leading cause mortality globally, Oceania, Eastern, Western, and Central Sub-Saharan Africa, and South Asia had the highest rates of air pollution, caused by differences in exposure (Roth, et al., 2020). Figure 2.4 compares the rankings of CVD disability-adjusted life years attributable to modifiable risk factors in 1990 and 2019.

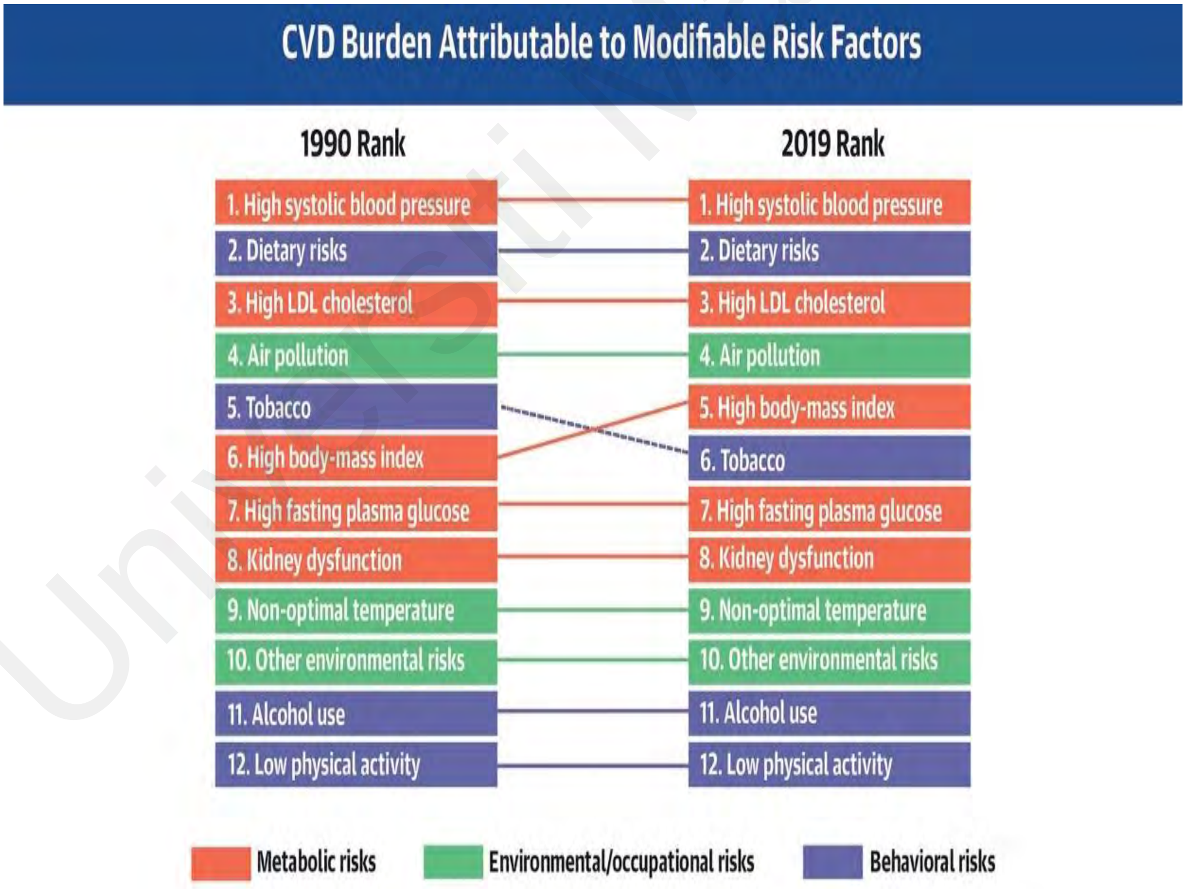

CVD Burden Attributable to Modifiable Risk Factors
1990 Rank
2019 Rank
1. High systolic blood pressure
2. Dietary risks
3. High LDL cholesterol
4. Air pollution
5. Tobacco
6. High body-mass index
7. High fasting plasma glucose
8. Kidney dysfunction
9. Non-optimal temperature
10. Other environmental risks
11. Alcohol use
12. Low physical activity
1. High systolic blood pressure
2. Dietary risks
3. High LDL cholesterol
4. Air pollution
5. High body-mass index
6. Tobacco
7. High fasting plasma glucose
8. Kidney dysfunction
9. Non-optimal temperature
10. Other environmental risks
11. Alcohol use
12. Low physical activity
Metabolic risks
Environmental/occupational risks
Behavioral risks

Figure 2.4: Comparison of CVD burden attributable to modifiable risk factors in 1990 and 2019, air Pollution as the leading environmental/occupational risk (Photo sourced from Roth G., 2018).

Non-modifiable risk factors, such as age, gender, and family history, also play a crucial role in the development of ACS (Varghese & Kumar, 2019). Ageing can induce changes in the heart and blood arteries that lead to hypertension, or high blood pressure. Studies discovered that older middle-aged age is a significant predictor of adverse cardiovascular events, such as myocardial infarction, stroke, and mortality in patients with ACS (Gillis, et. al., 2014; Soiza, et. al., 2005; Ahlgren, et. al., 1997). Furthermore, in Ranjith, et al. (2005) study found that hospital mortality was low in young and middle-aged patients but higher in older patients.

Research has been carried out by Kasim, et al. (2022) in Asian elderly patient using the ML and deep learning approach, and has identified that age, fasting blood glucose, heart rate, Killip class, oral hypoglycemic agent, systolic blood pressure and total cholesterol as common predictors of mortality in the elderly.

In the 7th National Cardiovascular Disease – Acute Coronary Syndrome (NCVD-ACS) Report, the general risk factors for ACS focusing on the Malaysia cohort includes patient's demographics, status before event, clinical presentation and examination, ACS onset details, baseline investigation, electrocardiography (ECG), clinical diagnosis at admission, invasive therapeutic procedures, and pharmacological therapy (Ahmad., 2022). Several studies have shown that the risk factors for ACS in the western cohort population show different cardiovascular profiles, highlighting the need for tailored prevention and treatment strategies based on geographic and cultural differences (Ueshima, et al., 2008; Asia Pacific Cohort Studies, 2005; Natarajan, 2018; Ohira & Iso, 2013; Nag & Ghosh, 2013). Hence, recognizing that the risk factors for ACS may vary across populations is essential for determining the most effective prevention and treatment strategies.

According to the results released by the National Health and Morbidity Survey (NHMS) (2020), which was conducted by the Institute for Public Health (IPH) in Malaysia to determine the prevalence of non-communicable diseases, risk factors for non-communicable diseases, healthcare demand, and health literacy levels in the country. The findings of the survey indicated that CVDs, such as stroke and coronary heart diseases, are the leading causes of death in Malaysia.

High blood sugar, high blood pressure, and high cholesterol are major risk factors for cardiovascular disease. The survey received a total of 14,965 responses, with a response rate of 87.2%. The key findings of the survey indicated that CVDs are the leading causes of death in Malaysia. According to the report, the three primary risk factors are diabetes, high blood pressure, and high cholesterol as shown in figure 2.5. According to the report, 1.7 million individuals in Malaysia presently live with all three of these major risk factors, and 3.4 million people in Malaysia currently live with two of these significant risk factors (IPH, 2020).

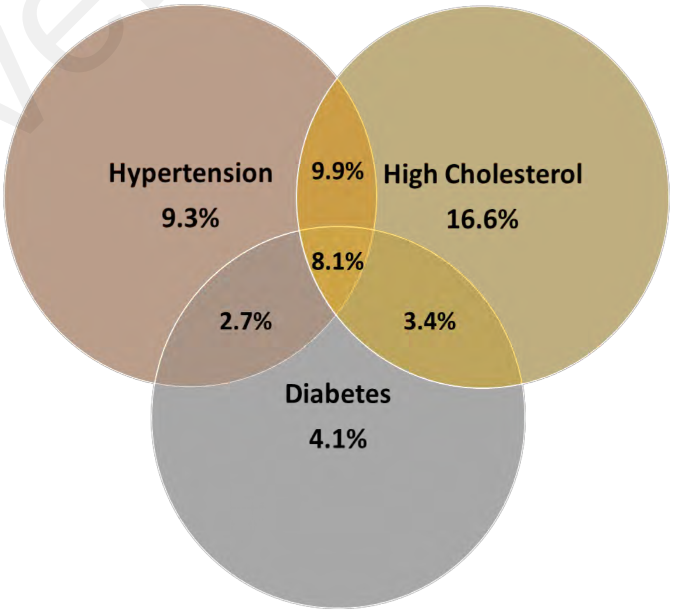


Figure 2.5: High Risk Factors in Malaysia Venn diagram (Photo sourced from Institute of Public Health (IPH), 2020).

Table 2.2 below summarizes existing studies on the common risk factors found in ACS patients.

Table 2.2: Previous research on the common risk factor found in ACS patients.

Authors	Studied Populations	Number of Populations	Risk Factor Findings
(Ralapanawa, et al., 2019)	Sri Lanka	300	Smoking, alcohol consumption, hypertension, diabetes Miletus, history of ACS, and dyslipidemia
(Kasim, et al., 2022)	Malaysia	3991	Age, fasting blood glucose, heart rate, Killip class, oral hypoglycemic agent, systolic blood pressure, and total cholesterol
(Sidhu, et al., 2020)	India	651	Age, gender, hypertension, diabetes mellitus, dyslipidemia, past history of ischemic heart disease, and smoking
(Martinez-Sanchez, et al., 2016)	Mexico	8296	Age, Gender, STEMI, previous heart failure, hypertension, smoking, dyslipidemia, previous angina, previous myocardial infarction, previous PCI and previous CABG
(Mirza, et al., 2018)	Iraq	100	Diabetes Mellitus, Hypertension, Smoking, Family history of ACS, Obesity, Number of diseased vessels.
(Lu & Nordin, 2013)	Malaysia	13591	Ethnicity, age, gender, BMI, smoking, diabetes mellitus, hypertension, dyslipidemia, family history of premature coronary artery disease
(Juhan, et al., 2019)	Malaysia	16673	Diabetes mellitus, hypertension, family history of CVD, renal disease, PCI, Killip class, and age
(Alhassan, et al., 2017)	Northern Saudi Arabia	156	Sex, Nationality, Age, Hypertension, Ischemia Heart Disease, Smoking, Family History of IHD, Family history of DM, Family History of Dyslipidemia

Table 2.2, continued

Authors	Studied Populations	Number of Populations	Risk Factor Findings
(Mansoor, et al., 2017)	United States	12047	Age, number of chronic conditions, coagulopathy, hypertension, renal failure, family history of CAD, angiography, PCI, dyslipidemia, CAD, smoking, cardiogenic shock
(Cheema, et al., 2020)	Pakistan	300	Sex, Dyslipidemia, Diabetes Mellitus, Hypertension, Family history for ACS, Smoking
(Esteban, et al., 2014)	Spain	123	Smoking, Hypertension, Diabetes Mellitus, Obesity, Prior dyslipidemia, Hypertriglyceridemia, Low HDL-cholesterol, Total cholesterol, LDL cholesterol, HDL cholesterol, Triglycerides
(Suzuki, et al., 2019)	Japan	29832	Age, gender, hypertension, dyslipidemia, diabetes, heart failure, ischemic heart disease, valvular heart disease, cardiomyopathy, atrial fibrillation.
(Ke, et al., 2022)	China	6482	Killip class, D-Dimer, NT-proBNP, LVEF, LDH, Diagnosis, CTnl, age, LDL, and HDL.
(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.
(Ahmad, et al., 2011)	Malaysia	525	Hypertension, diabetes, dyslipidemia, smoking history, previous history of CAD, family history of CAD
(Vernon, et al., 2019)	Australia	3081	Hypertension, diabetes mellitus, hypercholesterolemia, smoking, Killip class, cardiac arrest at admission, systolic blood pressure, and hospital transfer

ACS has also been linked to air pollution as a risk factor. According to existing studies, exposure to high levels of particulate matter, nitrogen dioxide (NO_2), and Sulphur dioxide (SO_2) in the air can increase the risk of ACS (Murad, 2012; Miller, et al., 2007). Li et al. (2017) discovered a significant correlation between long-term air pollution exposure and the occurrence of ACS, with each 10 g/m3 increase in PM2.5 concentration associated with a 12% increase in the risk of ACS.

In addition, an American study found that exposure to air pollution was associated with an increased risk of hospitalization for ACS, with the risk being highest for women and elderly adults (Brook, et al., 2010). These results emphasize the significance of minimizing air pollution as an approach to prevent ACS and other CVD. The focus of this study, where air pollution is one of the risk factors of ACS.

2.3 Air Pollution

Air pollution is the presence of one or more contaminants in the atmosphere by any chemical, physical, or biological agent, such as dust, fumes, gas, mist, odor, smoke, or vapor, in quantities and duration that are detrimental to human health and modify the natural characteristics of the atmosphere (WHO, 2022). In short, air pollution can be defined as a decrease in the air quality due to the presence and release of inorganic and organic pollutants into the environment (Mohamed & Awad, 2022).

Air pollution is a major environmental health concern, as it can have negative impacts on both the natural environment and human health (Manisalidis, et al., 2020). In 2019, 99% of the world's population resided in areas that did not meet WHO air quality guidelines. Approximately 89% of these premature fatalities occurred in low- and middle-income countries, with the majority occurring in the South-East Asia and Western Pacific WHO Regions (WHO, 2022). Common sources of air pollution such as household combustion devices, motor vehicles, industrial facilities, and forest fires. Pollutants that pose a significant threat to public health include particulate matter (PM), carbon monoxide (CO), ozone (O_3), nitrogen dioxide (NO_2), and sulphur dioxide (SO_2).

Nitrogen oxides (NOx) are a group of gases that are composed of nitrogen and oxygen. These gases are formed when fossil fuels are burned and are released into the air through various sources, including power plants and vehicles (Nirel & Dayan, 2001). NOx is known to cause cardiovascular and respiratory problems and are also harmful to the environment as they contribute to the formation of smog and acid rain (Gómez-García, et al., 2005). According to Zhang, et al. (2021), NOx would trigger increasing O_3 concentration, with sectoral emission control, the study demonstrates that in China could reduce more than 1.5-2% of emergency ACS hospitalizations for cardiovascular and respiratory diseases attributed to NOx and O_3 exposure.

Sulphur dioxide (SO_2) is a toxic gas that is produced when fossil fuels, such as coal and oil, are burned. It is released into the air through various sources, including power plants, factories, and vehicles (Joskow, et al., 1998). SO_2 is harmful to human health, as it can cause cardiorespiratory mortality and morbidity and it is also harmful to the environment as it contributes to the acidification of soil and water (Wu, et al., 2020; Wang, et al., 2018; Khaniabadi, et al., 2017; Khaniabadi, et al., 2017).

Ozone (O_3) is a highly reactive gas that is formed when sunlight reacts with pollutants in the air. O_3 is harmful to human health, as it can cause respiratory problems, and it is also harmful to the environment as it can damage crops and forests. O_3 is known as photochemical oxidant, and it was identified that the increased risk of heart failure is associated with photochemical oxidant level (Zhao, et al., 2016).

Particulate matter (PM) is a mixture of tiny particles and droplets that are suspended in the air. These particles and droplets can be composed of a variety of substances, including dust, dirt, soot, and other pollutants that can be inhale into human respiratory system which is harmful to human health, as it can lead to respiratory problems and various health issues, and it is detrimental to the environment as it can reduce visibility and cause other environmental problems (Fierro, 2000).

There are two types of PMs, PM10 refers to particulate matter that has a diameter of 10 micrometers or less, while PM2.5 refers to particulate matter that has a diameter of 2.5 micrometers or less. PM2.5 is of particular concern because it is small enough to enter deep into the lungs and potentially cause health problems (Feng, et al., 2016; Lall, et al., 2004; Lu, et al., 2019).

Addressing air pollution, which is the second highest risk factor for noncommunicable diseases, is key to protecting public health. An estimated 4.2 million deaths globally are linked to ambient air pollution, with 25% of deaths and diseases attributable to ischemic heart disease (WHO, 2022). Although levels have declined in high-income countries (HICs) over the past 25 years, they have risen sharply over that same period in China, India, and other low- and middle-income countries (LMICs), threatening public health and economic development (Stanaway, et al., 2018; Boogaard, et al., 2019). Therefore, urgent action is needed to reduce air pollution levels, particularly in LMICs where levels have risen sharply in recent years.

Malaysia is an industrialization-focused developing country. Furthermore, the preferences of using private cars are a common practice in Malaysia, resulting in haze and transboundary air pollution. Consequently, air pollution has become a significant problem in Malaysia in recent years. Air pollution, such as O_3 and airborne particles, has been linked to an increase in hospital admissions and mortality (Usmani, et al., 2020) and the short-term exposure of high-level air pollution often led to an acute condition (Afroz, et al., 2003). However, there is a significant research gap concerning the health effects of air pollution in Malaysia, concerning in the lack of comprehensive studies and data collection for environmental epidemiological analysis makes determining the full extent of health consequences from air pollution in the country difficult.

2.4 Impact of Air Pollution and Acute Coronary Syndrome (ACS) Onset

The impact of air pollution on the onset of ACS has been well documented in numerous studies. Historically, air pollution research has concentrated on adverse effects on the respiratory system (Pope CA & Dockery, 2006); however, numerous epidemiologic studies now link long-term exposure to air pollution with cardiovascular morbidity and mortality (Hoek, et al., 2013; (Brook, et al., 2010; O'Toole, et al., 2008).

Air pollution is a known risk factor for cardiovascular events, with both short term and long-term exposure associated with an increased risk of cardiovascular events (Yang, et al., 2019; Roth G. et al., 2018; Franklin, et al., 2015; Koulova & Frishman, 2014; Yamamoto, et al., 2014; Chuang, et al., 2011). Short term exposure to air pollution is typically considered to be exposure over a period of hours or days, while long term exposure is defined as exposure over a period of months or years.

ACS is the main acute presentation of IHD where there is significant myocardial ischemia leading to significant morbidity and mortality (Zhao, et al., 2016). Studies have shown that short term exposure to air pollution is associated with an increased risk of ACS events, with the risk being highest for those with pre-existing heart conditions (Chen, et al., 2022; Kuźma, et al., 2021; Gestro, et al., 2020). The effects of short-term exposure, particularly exposure to fine particulate matter air pollution, have the likelihood of triggering acute coronary events, especially patient with severely diseased coronary arteries (Pope III, et al., 2015).

A study by Dastoorpoor, et al. (2019), found that there is a significant increase in cardiovascular admission in the total population in presence of air pollution in the Middle East. The results show that PM10, NO_2, CO and SO_2 significantly increased cardiovascular cardiac hospitalizations. Similarly, a study by Yao, et al. (2020) stated that short term exposure to ambient air pollutants causes increase in health burden and economic loss in China, suggesting that adverse health affect due to short-term exposure of ambient air pollutant should not be neglected. In addition, the levels of NOx are positively correlated with the number of ACS hospitalization in the Valencia region (Ruvira, et al., 2023).

Long-term exposure to high levels of PM can increase the risk of respiratory and cardiovascular diseases (Chen & Hoek, 2020; Yuan, et al., 2019; Pelucchi, et al., 2009). While short-term exposure can result in acute health effects such as coughing, wheezing, and difficulty breathing, it also has a significant impact on hospitalizations and mortality (Shang et al., 2013; Bae, 2014; Li et al., 2016).

In addition to its role in triggering the onset of ACS and increasing hospital admissions among cardiovascular patients, high levels of air pollution have also been found to have a significant impact on the mortality of ACS patients. A study carried out by Bañeras, et al. (2018), found a positive association between short-term exposure to high levels of NO_2 and higher mortality in STEMI patients in Barcelona. Short-term exposures to PM2.5 and warm-season ozone were substantially associated with an increased risk of mortality in the US Medicare population from 2000 to 2012 (Di, et al., 2017). Likewise, the Chinese population exhibited significant associations between air pollution exposure and increased mortality risks (Shang, et al., 2013). In Malaysia, it is found that exposure to PM2.5 was associated with the increase premature mortality (Mazeli, et al., 2023).

While recent research has confirmed the short-term detrimental effects of air pollution on cardiovascular morbidity, there is a need for governments and policymakers to implement policies to reduce air pollution. However, previous studies have mostly used conventional statistical methods, such as general additive Poisson models, to identify the correlation between air pollution and ACS. The current study aims to enhance the existing studies by employing ML techniques to identify the significant air pollutants for hospitalization and mortality among ACS patients. Figure 2.6 below depicts the geographical map of death attributable to ambient air pollution in 2016.

Beverland, et al. (2012) compare the impact of short-term and long-term exposure to air pollution on mortality risk. The results showed that short-term exposure-mortality associations in cohort participants were of greater magnitude than in comparable general population time-series study analyses. Therefore, this study will focus on the short-term effects of air pollution exposure on ACS patients, as the studies described in the table below are designed based on a daily basis (time lag 0).

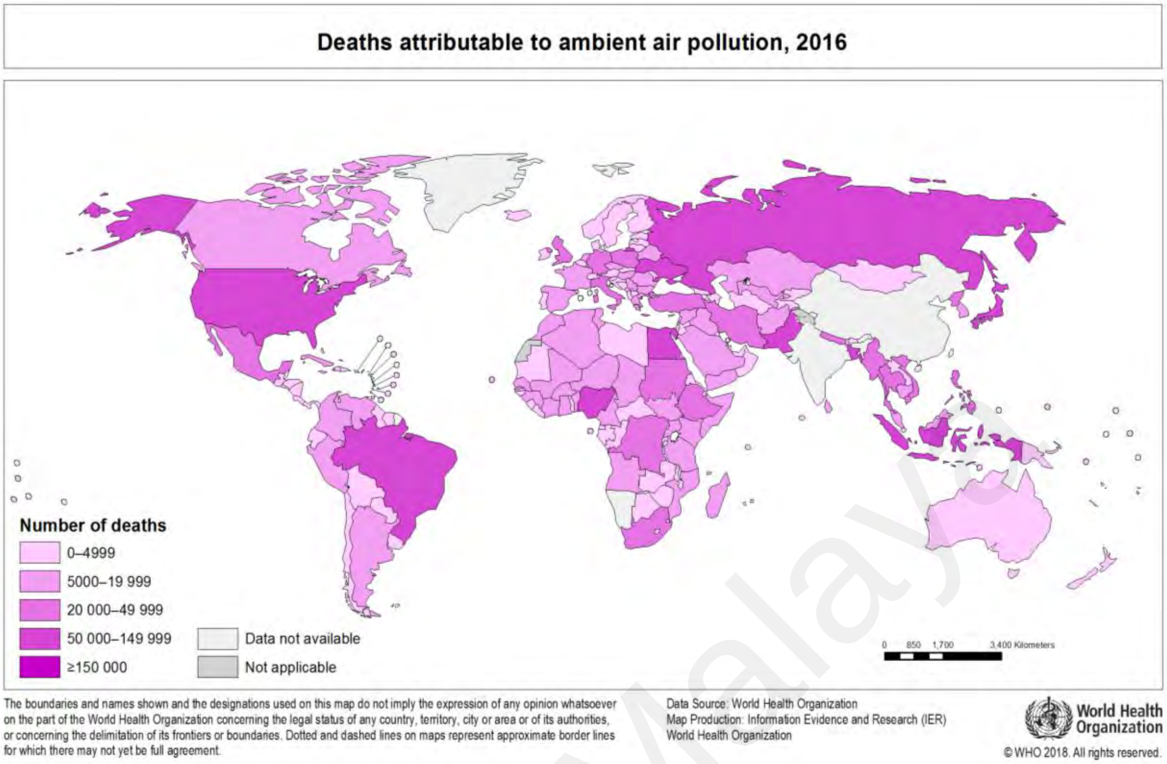

Deaths attributable to ambient air pollution, 2016
Number of deaths
0–4999
5000–19 999
20 000–49 999
50 000–149 999
≥150 000
Data not available
Not applicable
0 850 1,700 3,400 Kilometers
The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.
Data Source: World Health Organization
Map Production: Information Evidence and Research (IER)
World Health Organization
World Health Organization
© WHO 2018. All rights reserved.

Figure 2.6: Geospatial map illustrates the death attributable to ambient air pollution in 2016 (Photo sources from World Health Organization (WHO) 2022).

Table 2.3 summarizes studies on the short-term effects of air pollution on cardiovascular mortality, emphasizing the importance of studying deeper into the effects of short-term air pollution in the current study.

Table 2.3: Existing studies on short-term of air pollution on cardiovascular mortality.

Study (Reference)	Location	Study Design	Air Pollutants	Result
(Samet, et al., 2000)	United States	log-linear regression	PM10, O_3, SO_2, NO_2	Increase in PM10, the rate of cardiorespiratory fatalities increased by 0.68%
(Katsouyanni, et al., 2001)	Europe	Poisson regression	PM10, O_3, SO_2, NO_2	Increase in PM10, the rate of cardiorespiratory deaths by 0.6%. Slightly higher for elderly population. Increase in NO_2, the rate of mortality increases by 0.80%

Table 2.3, continued

Study (Reference)	Location	Study Design	Air Pollutants	Result
(Biggeri, et al., 2004)	Italy	Generalized linear model	PM10, O_3, SO_2, NO_2, CO	Increase in the rate of cardiovascular deaths of 0.40%for each elevation in NO_2, 0.93% in CO, 1.11% in SO_2, 0.54% in PM10
(Laden, et al., 2006)	United States	Cox proportional hazards regression	PM2.5	Increase in overall mortality associated with increase in PM2.5
(Bergmann, et al., 2020)	Worldwide	Literature Review Analysis	CO, O_3, SO_2, NO_2	Increase in SO_2, cause increment cardiovascular disease by 0.9828%
(Martins, et al., 2006)	Brazil	Generalized additive Poisson regressions	CO, PM10, O_3, NO_2, SO_2	PM10 and SO_2 increased congestive heart failure by 3.17% and overall cardiovascular illnesses by 0.89% at lag 0.
(Zhang, et al., 2017)	China	Generalized additive model	PM10, SO_2, NO_2	CVD mortality increased by 5.26%, 2.71%, and 0.68% with every SO_2, NO_2, and PM10 increase for lag 03 exposure.

2.5 Existing Research for CVD Hospitalization Rate and CVD Death Rate in the Presence of Air Pollution

In recent years, there has been an increasing concern regarding the impact that air pollution has on the health of the heart and circulatory system, specifically on the incidence of ACS. The association between air pollution and CVD has been the subject of research in several research, the majority of which have concentrated on mortality and hospitalization rates. Table 2.4 summarized existing studies on CVD hospitalization rate and mortality rate with the presence of air pollution, majority of studies have employed traditional statistical approaches. These conventional methods typically involve linear or logistic regression models, time-series analysis, or case-crossover designs. Most of the previous studies apply conventional statistical methods in understanding the association of air pollution and hospitalization rate, the mortality rate of CVD patients. Statistical techniques describe the relationship between variables based on possibility and statistical average. However, the reactions between air pollutants and influential factors are highly non-linear, leading to a very complex system of air pollutant formation mechanisms. Therefore, more advanced statistical learning (or ML) algorithms are usually necessary to account for a proper non-linear modelling of air contamination.

Table 2.4: Existing studies on CVD hospitalization rate and mortality rate summary.

Authors	Location	Air Pollutants	Methodology	Summary
(Tian, et al., 2019)	China	PM2.5	Quasi-Poisson regression Generalized Additive Model	Short-term exposure to PM2.5 can lead to an increase in hospital admissions for various types of cardiovascular disease, except for hemorrhagic stroke, even at levels of exposure that are within current regulatory limits.
(Phung, et al., 2016)	Vietnam	PM10, NO_2, SO_2, and O_3	Time-series regression analysis, Generalized Linear Model and Distributed Lag Model	PM10, NO_2 and SO_2 at lag-0 day shows significant association with cardiovascular admissions

Table 2.4, continued

Authors	Location	Air Pollutants	Methodology	Summary
(Liu, et al., 2019)	24 Countries	PM2.5 and PM10	Over dispersed generalized additive models with random-effects meta-analysis	Independent correlations between daily all-cause, cardiovascular, and respiratory mortality and short-term exposure to PM10 and PM2.5 in more than 600 cities worldwide.
(Chen R. Y., 2017)	China	PM2.5	Two-stage Bayesian hierarchical models Overdispersal Generalized Additive Models	Each 10-µg/m3 increase in 2-day moving average of PM2.5 concentrations was significantly associated with increments in mortality.
(Yitshak-Sade, et al., 2018)	New-England	PM2.5	Poisson regression	Impacts of short-term exposures to temperature, temperature fluctuation, and PM2.5. Compared to short-term exposures, long-term exposures to PM2.5 were associated with greater impacts.
(Wang, et al., 2020)	China	SO_2, NO_2, O_3, PM10	Time-stratified case-crossover design combining with distributed lag nonlinear model (DLNM)	The elderly aged over 65 years were susceptible to extreme pollution conditions.
(Zhao, et al., 2016)	Japan	PM2.5, O_3, NOx	Conditional logistic regression with lag 0, lag 1, lag 2 and lag 3	53006 emergency ambulance cases, studies shows that PM2.5 (lag3) had significant association with the incidents

Table 2.4, continued.

Authors	Location	Air Pollutants	Methodology	Summary
(Chang, et al., 2005)	Taiwan	PM10, NO_2, CO, O_3, SO_2, Temperature, Humidity	Odd Ratios and Conditional logistic regression.	Higher level of ambient pollutants increases the risk of hospital admission for CVD, primarily PM10 in all the pollutant models
(Soleimani, et al., 2019)	Iran	CO, O_3, SO_2, NO_2, and PM10	linear regression (GLM) and generalized additive model (GAM) estimating Poisson distribution.	Among the pollutants, CO, NO_2 and PM10 shows association with coronary artery disease hospital admission.
(Xu, et al., 2021)	China	PM2.5	Cox regression model	Long term exposure PM2.5 affect cardiovascular related mortality, especially towards CVD patients.
(Qiu, et al., 2020)	New England	PM2.5 and O_3	Generalized Inverse probability weighting and Linear Regression	PM2.5 had the potential to induce higher risk of CVD hospitalization
(Işsever, et al., 2005)	Istanbul	CO, SO_2, NO, NO_2 and PM10	Pearson correlation coefficient	Significant association between an increase in PM10 levels and the admission frequency for ACS.
(Dastoorpoor, et al., 2019)	Iran	O_3, SO_2, NO_2, CO, PM10 and NO	Quasi-Poisson regression	Significant increase in hospital admissions for cardiovascular diseases associated with various air pollutants

Table 2.4, continued.

Authors	Location	Air Pollutants	Methodology	Summary
(Ruvira, et al., 2023)	Spain	NO_2, NO, CO, SO_2, O_3 and PM	Mixed-effects model	ACS risk increase is related to high level of NOx and CO.
(Rajak & Chattopadhyay, 2020)	India	PM, PM2.5, PM10, SO_2, O_3, CO_2, CO, SPM and NO_2	Meta-analysis	Significant association between ambient air pollution exposure and increased premature mortality risk.
(Leem, et al., 2015)	Korea	PM2.5 and PM10	Epidemiology-based exposure-response functions	Air pollution was responsible for 15.9% of total mortality, or approximately 15,346 cases per year.

2.6 Machine Learning in Hospitalization and Mortality Prediction

Machine Learning (ML) has been widely incorporated in healthcare and clinical studies for discovering patterns from medical data sources and providing excellent capabilities to predict diseases (Shailaja, et al., 2018). ML has been used in several studies to predict hospitalization effectively across various medical conditions and settings. However, most existing studies focus on predicting hospitalization rates, with fewer addressing the prediction of mortality rates.

Additionally, there is lack of literature in predicting the hospitalization and mortality rates focusing on ACS onset due to air pollution. Further research is necessary, considering the potential health implications of air pollution on ACS incidents. Table 25 presents the summary of existing literature on ML predictions for hospitalization and mortality.

Table 2.5: Summary of existing literature on ML predictions for hospitalization and mortality.

References	Prediction Focus	Features Used	Methodology	Key Findings
(Usmani, et al., 2021)	Cardiorespiratory Hospitalization due to air pollution	CO, O_3, PM10, NOx, NO_2, NO, SO_2	Time Series ML Algorithms - ELSTM, LSTM, DL, Vector Autoregressive (VAR)	ELSTM model accurately predict cardiorespiratory hospitalization based on air pollution (RMSE: 0.002).
(Qiu, et al., 2020)	Peak demand days of cardiovascular diseases admissions	temperature, relative humidity, rainfall, PM2.5, PM10, SO_2, NO_2, CO and O_3	LR, SVM, ANN, RF, XGBoost, LightGBM.	LightGBM achieved the highest AUC (0.940) and other optimal metrics.
(Ravindra, et al., 2023)	Acute respiratory infections hospitalization on outpatients visits due to air pollution	PM2.5, NO, NO_2, NOx, NH_3, SO_2, CO, Ozone, Toluene, Eth-Benzene, Xylene, Benzene, MP-Xylene, RH, WS, WD, SR, AT, RF, Year, Week, Day, Month	RF, K-Nearest Neighbors, Linear model, LASSO, Decision Tree, SVM, XGBoost and Deep Neural Network with 5-layers	RF model outperforms the studied eight ML models with R^2 = 0.606.
(Miranda, et al., 2021)	COVID-19 Hospitalization	COVID-19 patients' medical history and self-reported symptoms	Decision trees, neural networks, SVM.	ML models achieve accuracies between 79.1% to 84.7%

Table 2.5, continued.

References	Prediction Focus	Features Used	Methodology	Key Findings
(Goto, et al., 2018)	Emergency Department disposition hospitalization due to Chronic Obstructive Pulmonary Disease and asthma	Demographics, Arrival mode, Vital signs, Chief complaint, Comorbidities	Lasso regression, RF, Boosting, Deep neural network, Traditional logistic regression.	Boosting provided the best prediction for critical care outcomes with C-statistics of 0.80. For hospitalization prediction, the random forest achieved the highest C-statistics of 0.83. Both outperformed the reference model – traditional logistic regression.
(Radović, et al., 2022)	Mortality rate for haemodialysis patients	Features from nephrology database	Kernel support vector machine, K-means clustering	The complete database predicted mortality 94.12% accuracy. When limited to the three most common diseases, accuracy was 96.77%.
(Xiao, 2021)	Mortality rate of reported COVID-19 patients.	Date of case report, age group, case demographic, hospitalization status, ICU admission, gender, race, ethnicity	Logistic Regression, Decision Tree, Neural Network, Light GBM	The Light GBM model had the best predictability among the evaluated models.

2.6 Mortality Risk Prediction for ACS

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 is a crucial component of the management of patients with ACS because it enables healthcare professionals to identify patients at high risk of adverse outcomes and implement appropriate interventions to prevent or manage these outcomes (Lee, et al., 2015). It is a useful instrument for ensuring that the intensity of preventive therapies corresponds to the level of absolute risk since ACS is a life-threatening condition associated with high morbidity and mortality (Yatsuya, 2018). Early identification of patients at high risk of mortality allows clinicians to initiate aggressive treatment strategies, including timely revascularization, pharmacological therapies, and management of comorbidities.

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, 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 Jr, et al., 2014). Several studies have found that there is a range of prediction models for various CVD outcomes (Wessler, et al., 2015; Beswick, et al., 2011; Matheny, et al., 2011). According to more recent assessments, the number of published prediction models has risen substantially since then.

The development of risk scores typically involves a combination of these methods. The initial selection of predictors is often done through statistical analysis, and ML techniques are then used to refine the model, using feature selection and identify the most important predictors. The resulting risk score is then validated using independent datasets to ensure its accuracy and generalizability (Aziida, et al., 2021).

The mortality prediction models developed utilized statistical methods and ML in clinical studies. These models integrate various clinical, demographic, and laboratory variables to provide an accurate estimate of the patient's risk of adverse outcomes. Available studies regarding mortality risk score in ACS do not consider as a factor in causing mortality in ACS patient. Although, previous studies have shown that air pollution exposure is associated with higher mortality rates in ACS patients. Therefore, there is a growing interest in developing accurate prediction models for ACS patient mortality based on air pollution exposure.

Examples of mortality prediction models for ACS include the Global Registry of Acute Coronary Events (GRACE) score, GRACE2.0 and the Thrombolysis in Myocardial Infarction (TIMI) score. The mortality prediction models (commonly known as risk scores) for patients with ACS are discussed further in subchapter 2.7 below.

2.7 Risk Scores for ACS

Risk scores for predicting mortality in ACS patients were derived from clinical trials and developed using large-scale clinical studies that involve collecting data on a range of patient demographics, medical history, and clinical presentation. According to Bawamia, et al. (2013), stated that patients with ACS are required to be risk-stratified so deliver the most appropriate therapy.

Risk scores are helpful tools for assessing the risk of ACS patients and allow accurate estimations of ischemic and bleeding risk for individual patients (Bueno & Fernández-Avilés, 2012). There are several risk scores used to assess the risk and mortality of ACS. The GRACE ACS Risk and Mortality Calculator estimates admission-6-month mortality for patients with ACS, and TIMI Risk Score estimates mortality for patients with ST-segment elevation myocardial infarction (STEMI) and non-ST-segment elevation myocardial infarction (NSTEMI).

2.7.1 Thrombolysis in Myocardial Infraction (TIMI)

The TIMI (Thrombolysis in Myocardial Infarction) risk score is a widely used risk stratification tool for patients with ACS. The TIMI risk score has distinct models that have been developed for the two major subtypes of ACS, namely STEMI and NSTEMI.

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 (Morrow, et al., 2000). The higher the point obtained indicated the higher risk which was determined based on 8 clinical risk indicators with the possible points around 14. The point for each variable is shown in Table 2.6.

Table 2.6: The variables and points for TIMI for STEMI.

TIMI (STEMI) 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
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 risk, 3 to 5 as intermediate risk, and >5 as high risk (Correia, et al., 2014).

TIMI risk score for NSTEMI/UA used the TIMI 11B clinical trial for the composite endpoint of mortality at 14 days in the year 2000 (Antman, et al., 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, known as 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 (Rao & Agasthi, 2023). The point for each of the variables for patients with NSTEMI/UA is shown in Table 2.7.

Table 2.7: The list of variables and points for TIMI risk score for NSTEMI/UA patients.

TIMI (NSTEMI/UA) Score Variables	Point
Age >=65 years	1
At least 3 risk factors for CAD (family history of CAD, hypertension, hypercholesteremia, diabetes or being a current smoker)	1
Significant coronary stenosis (prior coronary stenosis >50%)	1
ST deviation	1
Severe anginal symptoms (>= 2 anginal events in last 24 h)	1
Use of aspirin in last 7 days	1
Elevated serum cardiac biomarkers	1

Morrow et al. (2001) proposed that the TIMI risk score was very suitable in developing countries because it has a low-cost risk estimation. It was developed in a clinical trial population. However, it mainly derived from a Western cohort with less participation from the non-western population.

The TIMI risk score is a bedside tool that is easy to calculate and provides a means of risk stratification for patients with ACS. TIMI risk score is also an effective risk stratification tool for patients with potential ACS in the emergency department (Khan, et al., 2022; Graham, et al., 2013; Hess, et al., 2010). Study by Selvarajah, et al. (2012) conducted a study about the validation of TIMI risk score for STEMI in the Asian population, comparing to the TIMI population, the study population are younger and had more complications, and TIMI risk score are applicable for Asian population and can be used for risk stratification of STEMI patients.

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, however the respondents felt TIMI lacked crucial risk factors for clinical decision-making in older individuals (>= 75 years) with ACS, such as non-traditional cardiovascular risk factors. 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 and accuracy.

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.

2.7.2 GRACE

The Global Registry of Acute Coronary Events (GRACE) risk score is another widely used tool for predicting mortality in patients with ACS. The GRACE risk score was developed in 2006 based on a multinational registry of 43,810 patients with ACS that comprises 94 hospitals in 14 countries recruited in the global registry of acute coronary events (GRACE) research between April 1999 and September 2005 and has since been validated in multiple independent cohorts (Fox, et al., 2006).

The objective of GRACE risk score is to help assess and manage ACS patients by predicting the cumulative six-month risk of death or myocardial infarction using easily identifiable characteristics. The score includes eight variables, including age, heart rate, systolic blood pressure, serum creatinine level, Killip class, cardiac arrest upon admission, ST-segment deviation, and elevated cardiac enzyme levels. Each variable is assigned a weighted point score, and the sum of the points determines the patient's risk of mortality. The simplified model proved reliable, with prospectively verified C-statistics of 0.81 for in-hospital patients' mortality and 0.73 for death or myocardial infarction from admission to six months following discharge (Fox, et al., 2006).

While there are 9 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 percutaneous coronary intervention (PCI). The table 2.8 below shows the comparison between the two prediction time points, where 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 2.8: The GRACE score comparison of variables between the two points predictions.

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

It has been shown that the GRACE mortality risk score has excellent predictive value for mortality and adverse cardiovascular outcomes in patients with ACS (Boukerche, et al., 2023; Chen, et al., 2022; Neves, et al., 2021; Pieper, et al., 2009). In addition, the GRACE score for predicting long-term mortality still maintains its outstanding performance in predicting long-term mortality for NSTEMI/UA patients (Bouzas Cruz, et al., 2021).

According to Chen, et al. (2018), GRACE risk score was more accurate and have good discriminatory accuracy in predicting long-tern mortality when compared to TIMI risk score due to stratification by the tertials of GRACE provided more prognostic information than the TIMI risk assessment. Besides, GRACE is easier to conduct and use. Shuvy, et al. 2018) proposed that the GRACE risk score reduced the mortality rate of ACS patient significantly due to its ability to stratify the patient.

Although GRACE was developed on a multinational registry, these scores were derived and validated in predominantly Caucasian populations (Bulluck, et al., 2019; Authors/Task Force Members, et al., 2012; Morrow, et al., 2000). The performance of the GRACE score in predicting all-cause death at 6 months was poor in Kao, et al. (2020) study, most probably because of the population study was ACS patients with diabetes, thus, applying the GRACE risk score which does not consider diabetes in the scoring system, proving it less effective in predicting the ACS patient's mortality. Chan, et al. (2011) suggested that recalibration of the GRACE score may significantly enhanced risk estimation and may facilitate the adaptation of externally developed risk scores to Asian practice. Nevertheless, its clinical application in Malaysia is lacking (Sallehuddin, et al., 2017).

Overall, the GRACE risk score is a useful tool in predicting mortality and adverse cardiovascular outcomes in patients with ACS has recently received a class IIa recommendation in European guidelines (Collet, et al., 2021), but its accuracy may vary depending on the patient population and the specific outcome being predicted. Table 2.9 shows the GRACE Score interpretation.

Table 2.9: GRACE risk score mortality risk interpretation.

GRACE Score Range	Mortality Risk
0-87	0-2%
88-128	3-10%
129-149	10-20%
150-173	20-30%
174-182	40%
183-190	50%
191-199	60%
200-207	70%
208-218	80%
219-284	90%
>=285	99%

2.7.3 GRACE2.0

GRACE 2.0, developed by Fox et al. (2006), is an improved and externally validated version of the GRACE score for predicting ACS outcomes over a longer term. The previous GRACE version, introduced in 2000, had limited predictive ability up to only 6 months.

GRACE 2.0 features improved discrimination and practicality based on linear associations and has been validated over a longer term of 1 to 3 years with substitutions possible for creatinine values and Killip class, performing almost as well (Fox, et al., 2006). The main difference between GRACE and GRACE2.0 is GRACE2.0 allows for substitutions of Killip Class for diuretic usage and for serum creatinine with history of renal dysfunction.

GRACE 2.0 utilizes values obtained from β coefficients of regression models using non-linear functions from 32,037 patients in the GRACE registry, which were validated in the French registry of Acute ST-elevation and non-ST-elevation MI (FAST-MI) 2005. It is designed for use in acute and emergency clinical settings as well as electronic devices over an extended period, with data entry taking approximately 30 seconds. The values are summed up to provide an estimate of the probability of adverse outcomes without conversion to a point system (Eggers, et al., 2021).

A validation study of GRACE2.0 for patients with ACS in-hospital mortality in Canada showed that it discriminates well in all patient groupings and accurately predicts adverse results in ACS patients across Canada (Elbarouni, et al., 2009). Besides, a study conducted based on the Vietnamese cohort, stated that GRACE2.0 has a better performance in predicting 1-year post discharged mortality with AUC=0.703 ($p<0.001$) (Nguyen, et al., 2021). Similarly, in Akyuz, et al. (2016) study, where GRACE 2.0 exhibits AUC=0.77 for 1 year mortality risk assessment. Overall, study by Huang, et al. (2016) validate its performance in a contemporary multiracial ACS cohort showed strong model discrimination across ACS types and racial/ethnic subgroups and may be useful for normal clinical management of ACS patients.

However, the anticipated chance of in-hospital mortality may need to be adjusted based on the health care context and therapy advances, additional factors may influence outcome, especially in geographical populations and healthcare systems not evaluated in the multinational GRACE program (Ono, et al., 2021). Furthermore, there are limited studies and evidence carried out regarding the application of GRACE2.0 risk score for the Malaysian population (Ismail, et al., 2022).

2.7.4 Comparison of Risk Scores

Table 2.10 below gives an overview of the risk score mention in the subchapter 2.7.1 to 2.7.3

Table 2.10: Overview of the risk score mentioned in this study.

Features	Risk Scores			
	TIMI (STEMI)	TIME (NSTEMI)	GRACE	GRACE 2.0
Age	•	•	•	•
Past Medical History	•	•		
Risk factors		•		
Medication used		•		
CSS/Killip class	•		•	•
Signs and symptoms		•		
Cardiac arrest upon admission			•	•
Heart Rate	•		•	•
Systolic Blood Pressure	•		•	•
Weight	•			
ECG Findings	•	•	•	•
Cardiac enzymes		•	•	•
Creatinine level			•	•
Treatment time	•			
Possible range of scores	0 – 14	0 – 7	1 – 372	1 – 336
Cut-off of high risk	>= 4	>= 3	>= 140	>= 126

2.7.5 Limitation of Risk Scores

Risk scores are widely used tools in clinical practice to aid decision-making for patient management. However, they also have several limitations that must be considered. One limitation is that risk scores may not be generalizable to all patient populations, particularly those with unique demographic or clinical characteristics. The TIMI and GRACE risk scores were primarily derived and validated from the Western Cohort mainly Caucasian population and may underestimate some multiethnic Asian population (Sia, et al., 2022). Hence, the risk scores may not reflect the region's diversity and maybe only applicable to specific populations, and it may lead to inaccurate risk stratification and inappropriate treatment decisions (Peng, et al., 2017). Exclusion of the high-risk patients is also another limitation of the risk scores (Chen, et al., 2018). Saar, et al. (2018) assessed the risk-treatment paradox in NSTEMI patients according to the estimated risk by GRACE score and came with a conclusion that NSTEMI patients with higher risk receive less guideline, which linked to a worse prognosis. Van der Sangen, et al. (2022) also stated that optimal care was linked to lower mortality in intermediate-risk and high-risk patients but was less likely to be provided as mortality risk increased.

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). According to Kwon, et al. (2019), TIMI and GRACE risk scores lack a weight for the risk factors as they only consist of the major prognostic factors, important information may not be included. Nevertheless, risk models contain numerous independent variables, which limit their utility. Several simple cardiovascular risk scores have been proposed as alternatives (Huang, et al., 2021).

Although the risk scores models have been widely studied and validated, the accuracy of a risk score may also be affected by the quality and completeness of data used to develop and validate the score. Recent concern is that some of the risk stratifications were built 20 years ago using randomized controlled trials (RCT) data prior to the introduction of drug-eluting stents and newer generation antiplatelets, that may not accurately reflect the complexity of disease processes and individual patient characteristics (Kwon, et al., 2019).

Additionally, risk scores may not account for changes in clinical practice or treatment modalities over time, which may impact their accuracy in predicting outcomes. In Van der Sangen, et al. (2022) research highlighted that the GRACE risk score overestimated the absolute in-hospital and 1-year mortality risk in contemporary patients. 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).

Another limitation is that risk scores may not capture all relevant factors that contribute to a patient's risk of adverse outcomes, 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 (Peng, et al., 2017; Obermeyer & Emanuel, 2016). In addition, risk scores may not account for environmental factors such as air pollution exposure, which has been shown to be associated with cardiovascular disease risk. This is concerning given the increasing evidence linking air pollution exposure to increased mortality in ACS patients (Brook, et al., 2010). The exclusion of air pollution variables from risk scores may lead to an underestimation of the true mortality risk for patients in areas with high levels of air pollution. Therefore, there is a need to explore the impact of air pollution on ACS patients' mortality and to develop new risk scores that consider the effect of air pollution.

Recent studies have demonstrated the potential of ML techniques to construct mortality risk prediction models for ACS patients. ML algorithms can identify patterns and relationships in large datasets, resulting in accurate risk assessments (Wang, et al., 2021). For instance, a study by Shouval, et al. (2017) reported that an ML-based risk score outperformed traditional risk scores in predicting the risk of 30-day mortality in STEMI patients. The study applied 6 different ML algorithm, where RF achieves the highest AUC of 0.91, compare to traditional risk score, TIMI and GRACE, 0.82 and 0.87 respectively, demonstrating the viability and efficacy of ML tools for predictive modelling in cardiology's complex data scenarios will help clinicians develop tools for more precise patient risk stratification in presence of air pollution, which is the aim of the study. In section 2.8 will further be discussed about the application of ML.

2.8 Artificial Intelligence (AI)

There are numerous applications of artificial intelligence (AI) in the medical field. Algorithms and techniques based on artificial intelligence may assist in predicting health issues, assessing organ health, and preventing health hazards (Swapna, et al., 2022). AI in healthcare utilizes massive amounts of data for analysis and interpretation to support medical professionals in making faster, more accurate decisions (Bennett & Hauser, 2013).

In clinical research, historical electronic health records (EHRs) are used to create AI models that predict patient outcomes. EHR data used for AI models comprises patient demographics, health indices, medical conditions, biomedical pictures, and clinical notes, however organized medical claims data are rarely employed. Medical claims data may not accurately reflect patient health problems, but it does show patient health care access frequency and disease prevention or treatment involvement, which affects patient health outcomes (Tran, et al., 2021).

The mortality rate of individuals with CVD has been predicted using a variety of algorithms and predictor variables (Kasim S, et al., 2022a; Kasim S, et al., 2022b; Aziida, et al., 2021; Sastoeldraijer, et al., 2013). In Tran, et al. (2012) research, compared various AI architectures for predicting the mortality rate of patients with CVD using structured medical claims data, which could help health professionals choose AI models to accurately predict mortality among patients with CVD using only claims data prior to a clinic visit. In addition to predicting health issues and assessing organ health, AI prediction, classification, and regression algorithms assist the medical sector minimize health risks (Braun, et al, 2020; Swapna, et al., 2022).

AI approaches can overcome the constraints of conventional CVD incidence prediction models, that are used to generate conventional risk assessments. AI techniques such as ML and deep learning (DL) could contribute to improving cardiovascular care by simplifying precision cardiovascular studies (Krittanawong et al., 2017). Table 2.11 presents the comparison between conventional risk score and AI based risk prediction.

Table 2.11: Comparison between conventional risk scores and AI based prediction models.

Descriptions	Conventional Risk Score	AI-based Risk Prediction
Hypothesis	Yes	No
Approach	Estimates and explain data	Practical prediction from data
Measurement	Goodness-of-fit, coefficients	Accuracy, root mean square error, mean absolute error, area under the curve, precision, recall, etc.
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
Accuracy	Provide reasonable estimate of risk, but limit predictive ability in certain cohort	High accuracy in predicting risk in various patient populations, and can identify new risk factors and associations
Output	Validate the hypothesis, causality	Predict new data, identify new patterns
Adaptability	Require update periodically for changing risk factors and patient population	Able to adapt to new data and variables in real-time, allowing for continuous improvement in risk prediction
Limitation	Low data dimensionality and require assumptions, may not capture all relevant risk factors or interactions between features	Overfitting, data privacy, security issues and require large amounts of high-quality data
Risk factors	Clinical and demographic factors only	Multimodality

2.8.1 Machine Learning (ML)

Machine learning (ML) is an emerging technology that utilizes computational statistics to discover optimized algorithms, which can learn from and make predictions based on data (Özen, et al., 2009). ML methods have the necessary flexibility to construct classifiers with good predictive performance compared with the statistical approach in cardiovascular related mortality prediction. It functioned by recognizing a certain pattern from train data to construct good and accurate assumption and prediction. It can be considered as part of artificial intelligence because it can learn and adapt to the changing environment, providing solutions for all possible situations.

ML is a subset of AI that focuses on the development of algorithms that enable computers to learn from data without being explicitly programmed. It involves the use of statistical techniques to discover patterns in large datasets and make predictions or decisions based on those patterns (Edgar & Manz, 2017). Thus, developing an ML model with the best algorithm is vital to assist clinicians and the public to be aware of the presence of air pollution. For instance, these ML algorithms have been applied to overcome non-linear limitations and uncertainties to achieve better prediction accuracy.

ML comprises of automatic feature selection that enables manipulation of large numbers of predictors and does not require underlying assumptions regarding the relationship between input features and output. In the diagnosis of heart disease, ML approaches help to improve data-driven decision-making (Ahsan & Siddique, 2022). ML methods such as Linear Regression, Logistic Regression, Support Vector Machine (SVM), Random Forest (RF), Extreme Gradient Boosting (XGBoost) and stacked Ensemble Learning (EL) has been successfully applied to predict the occurrence of several clinical diseases, such as myocardial infarction, and the risk of mortality in previous studies (Kasim S, et al., 2023; Kasim S, et al., 2022a; Kasim S, et al., 2022b; Aziida, et al., 2021; Aziz F, et al., 2019; Peng, et al., 2017; Wallert, et al., 2017; Kim, 2017; Shouval, et al., 2017). Besides, the use of EL in the prediction of coronary artery disease and ACS has grown in popularity as a result of the substantial advancements in ML (Kasim S, et al., 2023; Zheng, et al., 2021; Sherazi, et al., 2021; Jamthikar, et al., 2021).

ML algorithms can be classified in general into two primary categories: supervised learning and unsupervised learning. The primary focus of our study concentrated on supervised learning, an ML model where algorithms are trained using labeled data. This approach ensures that the model can make predictions based on the input-output pairs, making it particularly suitable for tasks where the outcome variable is known.

(a) Supervised learning

Supervised learning, as it was name as “supervised” because the learning process is done under the seen label of observation variables (Wang, et al., 2021). Supervised learning was used to train the model based on sample dataset by giving that targeted output provided. In supervised learning, the algorithm is trained using a labelled dataset in which the inputs are annotated with the desired results. The computer then applies these labelled data to new, unlabeled data to create predictions or assign categories (Talabis, et al., 2015).

Regression and classification are two common techniques of supervised learning algorithms used in ML, where the regression is used to predict continuous output variables while the latter is used to predict categorical output variables (Wang, et al., 2021). In this study, prediction models are developed that are able to forecast hospitalization rate and mortality rate of ACS patients based on air pollution. Classification is used to predict a discrete value, which has been applied in predicting the mortality of ACS patients with selected ACS features in the presence of air pollution and forecast the patient's mortality risk.

The sample dataset will be divided into a training dataset and testing dataset whereby the training dataset was annotated whereas testing dataset was not annotated. Features and annotations in the training set are used to predict the outcome in the testing set in a model. However, the targeted output was provided to compare with the predicted output to increase the accuracy of prediction. If the result was not satisfied, the model is going to train again (Fabris, et al., 2017). RF, SVM, decision tree, logistic regression, k-nearest neighbour (KNN), gradient boosting are examples of supervised learning (Belyadi & Haghighat, 2021). Both regression and classification problems are solved using ML in this study.

(b) Unsupervised Learning

Unsupervised learning is used to detect naturally occurring patterns or groupings in data (Kohonen et al., 2001). This is a challenging task to evaluate, and the utility of unsupervised learning groups is frequently determined by their performance in subsequent supervised learning tasks. When the instances are unlabeled, these algorithms attempt to apply techniques to the input data to mine for rules, find patterns, summarize, and aggregate the data points, assisting in extracting useful insight and better communicating the data to the user. The Self-Organizing Map (SOM) is a well-known unsupervised learning method.

The following sections describe the ML algorithms that were used in this study.

2.8.1.1 Linear Regression

Linear regression is a supervised ML algorithm that models a target prediction value based on independent variables introduce by Galton (1886). It is a statistical method used to comprehend the relationship between two variables that are correlated linearly (Vij, 2023), involves fitting a line to the provided independent and dependent variables, the formula is given below:

$$y = mx + c \tag{2, 1}$$

The most fitted line is identified using the least squares, which minimizes the sum of squared differences between the predicted and actual values. Linear regression only applies to regression model as to predict the continuous outcome and forecast future trends (Gupta, 2023).

In early studies, linear regression has proven its usefulness in cardiovascular studies. In (Larsen, et al., 1993), the researchers utilized a linear regression model to predict survival rates among patients experiencing out-of-hospital cardiac arrest and demonstrated the model's effectiveness in facilitating the planning of community emergency medical services programs and systems. Besides, linear regression also applied in predicting in patients with heart failure mortality (Du, et al., 2023).

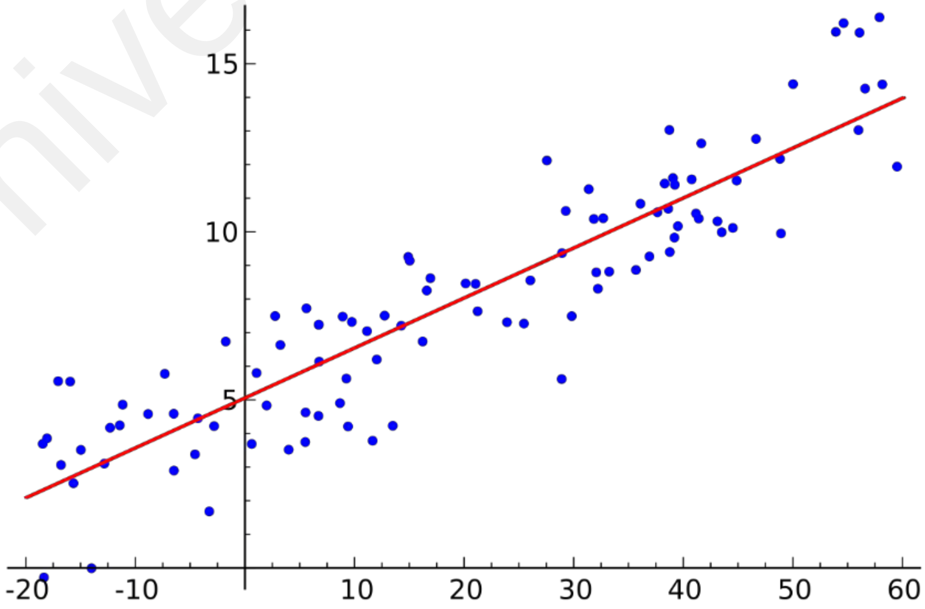


Figure 2.7: Linear regression graph (Photo sourced from Gandhi, 2018).

2.8.1.2 Logistic Regression

Logistic regression models are widely used in a variety of disciplines. It is being well known in the medical and health field with the notion of odd ratio applied to the studies such as smoking, cardiovascular disease, and other risk events (Hilbe, 2009). LR is a statistical method used for modelling a binary response variable by taking the value such 0 and 1 or yes and no. large sample sizes are required for LR to provide sufficient numbers in both categories of the response variable (Bewick, et al., 2005).

Assuming a Bernoulli distribution of the dependent outcome (y) that is conditional on a set of input predictors $(x_1, \dots, x_k)$ we can write $y \mid x_1, \dots, x_k \sim Bernoulli(p)$. Logistic regression (Cox, 1958) then estimates the binary response probability through the function as below:

$$\log\left[\frac{p_i}{(1-p_i}\right] = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \qquad (2,2)$$

where (β_0) is the intercept and $(\beta_1, \dots, \beta_k)$ are the estimated coefficients. Therefore, the probability of the suspected outcome can be obtained by the equation below:

$$p_i = \exp(\beta_0 + \cdots + \beta_i x_i)/(1 + \exp(\beta_0 + \cdots + \beta_i x_i) \qquad (2,3)$$

Usually, logistic regression used a cut-off at 0.5 to generate individual predictive probabilities for classification. Logistic regression lacked tuning parameters, which sets it apart from the other models.

In short, the logistic regression model was used to quantify the effect of a predictor in terms of a log odds ratio. The term can be explained from probability. Probability is the chance of an event likely to occur while odd means to the ratio of two probabilities. Therefore, odd ratio can describe an effect over the entire range of risk (Harrell, 2015).

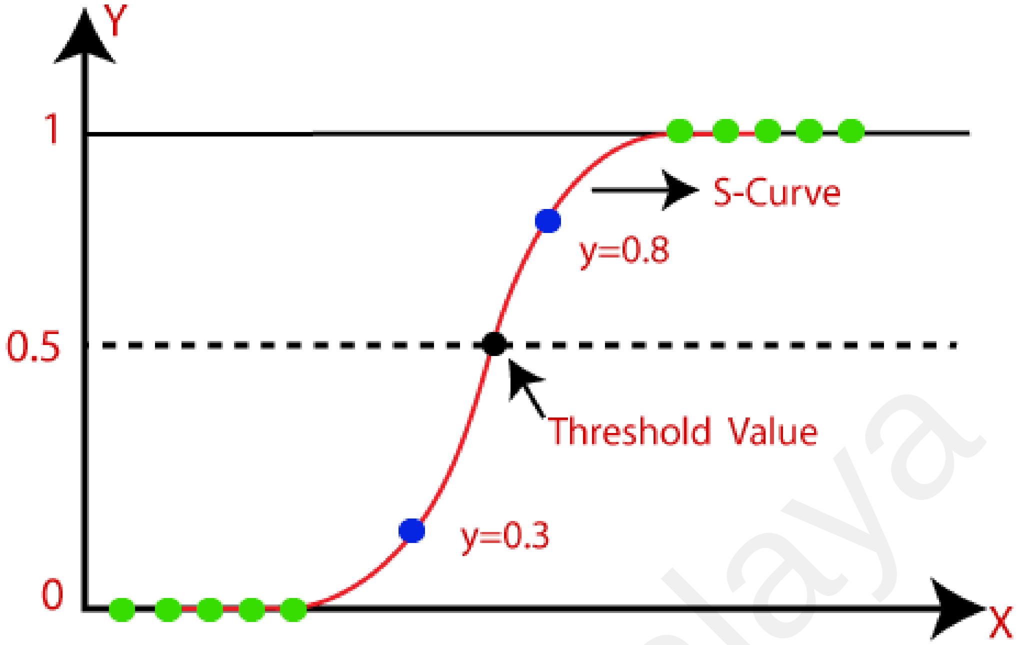


Figure 2.8: Logistic regression function to classify two maximum values (0 or 1) (Photo sourced from JavaTpoint, 2011).

2.8.1.3 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a common method of supervising ML (Vapnik, et al., 1995). SVM labelled each data point in n-dimensional space and classified them by a hyperplane. A hyperplane is a line which will generate an output that is divided into two classes well. A kernel function is used to transform non-separable data into linear separable data. It is very useful in SVM since it is required to separate the data in the n-dimensional space (Smola & Schölkopf, 2004).

In Battineni, et al. (2019) study, they proposed that introducing an optimal hyperplane is not easy because different hyperplane produced will influence its accuracy. The parameter like gamma and *C* value helped us to make a correlation between the hyperplane parameter to investigate better support-vector. Even in the different kernel function, they varied for them. It gives a chance to make an ideal hyperplane. He stated that SVM always chooses the optimal hyperplane with low gamma and high *C* value.

The proper parameter setting in the kernels will increase the classification accuracy. The main parameters in SVM are gamma and *C* value. The gamma parameter is used to define the distance for the single training data that can be reached. More far distance represented with low gamma value while close distance is represented by greater gamma value. The paper also showed that varying gamma value influences the model's performance, higher gamma value gives the better result. *C* value gives the trade-off training examples misclassification against decision surface simplicity. Higher *C* value gives an accurate result whereas lower c value ensures a smooth decision surface (Renukadevi & P., 2013).

SVM can classify and identify syndromes in coronary heart disease. Besides, it performed a better result compared to the decision tree algorithm with 82.5% accuracy against 80.4% accuracy (Chen, et al., 2007). Various studies also prove that SVM method exhibit good results and accurate in clinical diagnosis of cardiac disease (Gong & Wang, 2009; Alty, et al., 2003; Hongzong, et al., 2007). In Zhang, et al. (2012), stated that best SVM parameters with radial basis function show the highest classification accuracy in diagnosing coronary heart disease.

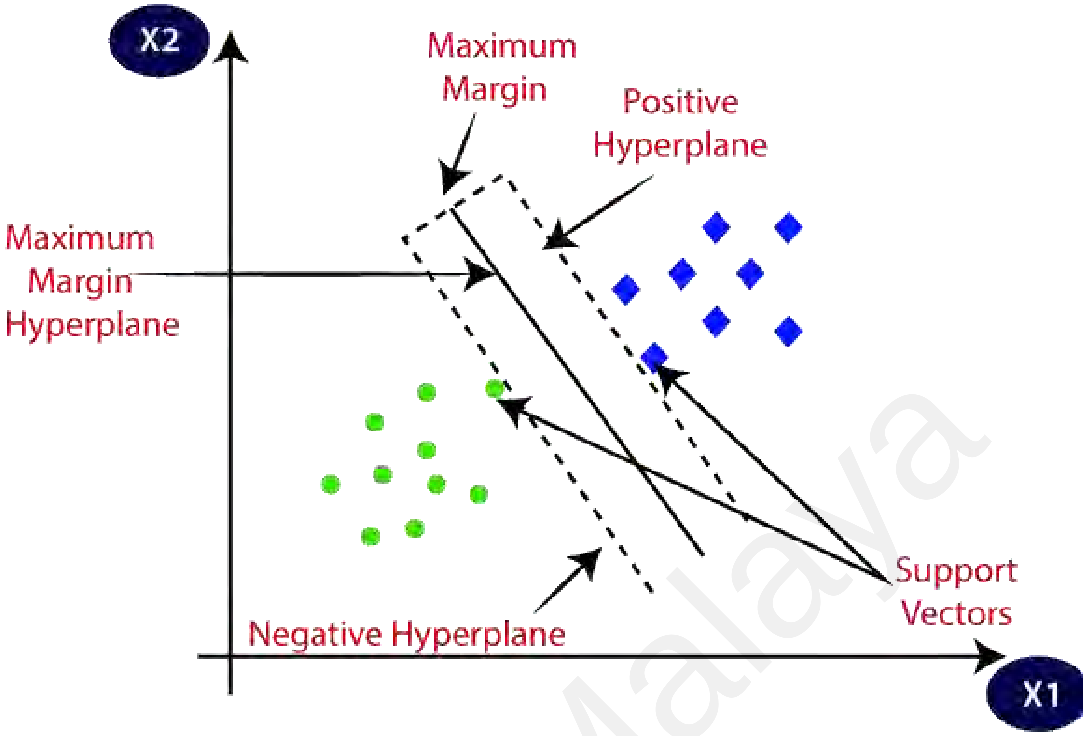


Figure 2.9: Support Vector Machine (SVM) separates two different categories that are classified using a decision boundary or hyperplane (Photo sourced from JavaTpoint, 2011).

2.8.1.4 Random Forest (RF)

According to Ho (1995), random forest (RF) is a method that will generate multiple decision trees based on the training data given. It splits the data into smaller and smaller trees, resulting in multiple trees and generates significant predictors that will influence the outcome. Nowadays, RF is a common method because it worked well in avoiding overfitting and increasing the accuracy in prediction.

RF is an ensemble method that builds decisions trees and incorporates the important predictors and their interaction during the learning process. Hence, there showed a rise in RF application in computational biology because it was nonparametric, interpretable, efficiency and accuracy for many types of data (Qi, 2012).

Sammut & Webb (2011) defined that RF is a hybrid of bagging algorithm and random subspace method. It used decision trees as the base classifier. Each tree is constructed from a bootstrap sample in the original dataset. The RF method was unpruned, therefore it avoided overfitting. Random subset method was used to identify the feature and the subset size is split at each branch in the tree to obtain the diversity of the classifiers. Both methods yielded low bias and high variance but low correlation trees. Combining the trees to achieve low bias and low variance forest.

Bootstrap aggregation can be short form as bagging. Breiman (2001) demonstrated that each tree was built based on random samples from the training set where replacement may occur, resulting in different trees. Hence, RF used bagging method to build large and not correlated trees and then average them. It draws a random subset of features for training the individual trees, resulting in better predictive performance and it is much simpler and easier to tune.

The parameters of random forest affect the result of the machine's prediction. Out-of-bag (OOB) errors are an estimate measuring prediction error in an RF. It can be affected by the parameters such as *mtry*, *ntree* and *nodesize*. OOB error is largely influenced by *mtry* but seems not readily affected in *ntree* and *nodesize*. Increasing *mtry* lead to small decreases in error rate. Conversely, decreases in *mtry* will lead to increases in error rate. *Ntree* gives more impact in the feature section. Larger *ntree* values will generate slightly more stable values of feature importance. However, *ntree* values is directly proportional to the time of execution. Change in *nodesize* values give a negligible effect (Díaz-Uriarte & De Andres, 2006).

The RF method gave a good performance and played an important role in the medical field. It was able to classify normal and congestive heart failure with 100% accuracy classification (Masetic & Subasi, 2016).

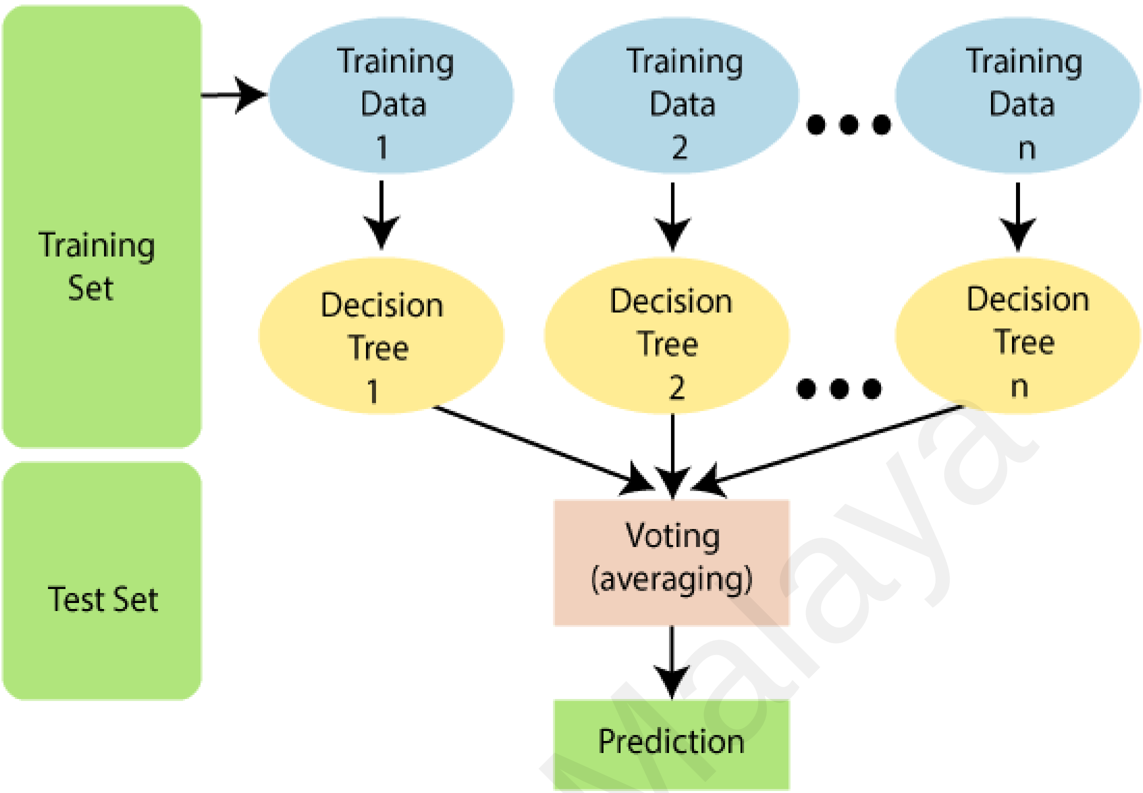


Figure 2.10: The architecture of random forest (RF) algorithm (Photo sourced from JavaTpoint, 2011).

2.8.1.5 Naïve Bayes (NB)

The Naïve Bayes (NB) algorithm was introduced and developed by Thomas Bayes, in 1968. It is based on Bayes' theorem (Bayes, 1968). In the 20th century, statisticians and computer scientists improved and popularized the technique (Wu, et al., 2008). NB is a probabilistic algorithm primarily used in supervised learning for classification problems. It is based on Bayes' theorem and the assumption of independence among the features.

The algorithm is based on Bayes' theorem, assuming that $|A| \neq 0$ and $|B| \neq 0$, which states that the probability of an event (A) given the occurrence of another event (B) can be calculated as the product of the probability of B given A and the probability of A, divided by the probability of B (Berrar, 2018). The formula for NB classification can be written as follows:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2,4}$$

Where;

- *P(A|B)* is the probability of event A occurring given that event B has occurred.
- *P(B|A)* is the probability of event B occurring given that event A has occurred.
- *P(A)* is the prior probability of event A occurring.
- *P(B)* is the prior probability of event B occurring.

NB classifiers use the probability theory to find the most likely classification of an unseen (unclassified) instance chooses the class with the highest probability as the prediction. The algorithm performs positively with categorical data, but it performs poorly if numerical data is present in the training set (Chen, et al., 2020; Vembandasamy, et al., 2015).

The strength of NB lies in its simplicity and efficiency, as it requires minimal computational resources and works well even with small datasets, which can be advantageous in situations where data may be limited (Saritas & Yasar, 2019). NB has the ability to handle high-dimensional data, and less prone to overfitting, which can be challenging for other algorithms (Bai, et al., 2023).

Vembandasamy, et al. (2015) and Maheswari & Pitchai (2019) applied the NB algorithm in developing heart disease prediction system, providing instance guidance for heart disease to the user based on the accurate result prediction. Besides, Khennou, et al. (2019) also utilized the NB algorithm as one of the techniques for predicting heart disease, along with SVM and decision-based systems. The proposed approach achieved an accuracy of 86% using the NB algorithm on the Heart Disease Dataset from the UCI machine learning library.

2.8.1.6 Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) is another ML algorithm. It is 10 times quicker than existing gradient boosting (Chen & Guestrin, 2016). This is because parallel distribution learning tree is used makes the learning process faster than other methods. Large data can be used in this method. In general method of boosting, predictors in each tree are achieved by weighting. The weighting to all independent variables of each tree will be assigned in each iteration. The weight played a role in predicting results. The weight will be increased in some predictors if any misclassification occurs. Next, the new individual tree will build on more weighting value. Tree is created in sequential form (Redpath & Lebart, 2005).

XGBoost is available as an open-source package. The main factor behind XGB making it a success is its scalability. The scalability of XGB is from several important system and algorithmic optimizations such as a novel tree learning algorithm which responsible for handling sparse data if and only if there has missing value in the dataset as well as a theoretically justified weighted quantile sketch which handling weights value in the tree. Besides, it runs with greedy algorithm for split finding. It means that it computed all the possible splits for continuous features and worked up with them to build the tree for machine learning. Then the weight value is used to meet the second criteria and find the best split points (Chen & Guestrin, 2016).

XGBoost is an implementation of gradient boosting. However, XGBoost gives a more accurate result because it used a more regularized form of gradient boosting which improves model generalization capabilities that can control overfitting. Besides, it used parallel tree learning makes the learning process faster. It is more capable of handling missing value compared to gradient boosting.

In the study by Li, et al. (2020), the authors applied the XGBoost algorithm to predict the likelihood of diabetes in patients. The results of the study showed that the XGB algorithm was able to achieve a prediction accuracy of above 80.2% when predicting the likelihood of diabetes in patients. The authors also compared the performance of XGBoost with other ML algorithms such as RF, Neural Network and SVM, and found that XGBoost outperformed these algorithms. Overall, the study provides strong evidence that ML algorithms can be effectively applied in the medical field to predict diseases like diabetes, cardiovascular disease etc.

2.8.1.7 Ensemble Learning (EL)

Ensemble Learning (EL) has gained much popularity in the ML community, due to its ability to improve the accuracy and robustness of predictive models. It combines multiple models to improve prediction accuracy and reduce the risk of overfitting (Park & Kim, 2021).

The three main EL methods are stacking, boosting, and bagging (Simske, 2019). Bootstrap aggregating, commonly known as bagging, was introduced by Breiman (1996). RF is an example of the bagging ensemble method, where decision trees are trained on N random subsets of the data, drawn with replacement. The predictions from each model are then combined to produce the final prediction. This technique involves training multiple models on randomly selected subsets of the data and is useful for reducing overfitting and increasing the accuracy of the model (Simske, 2019; Breiman L., 2001).

Another type of EL is boosting, which includes iteratively training models to rectify the errors made by prior models. This allows the system to make the best decision possible by considering the results from the samples in proportion to how well they contribute to the general accuracy of the system (Simske, 2019). AdaBoost is the most well-known example of EL boosting algorithm, where it trains weak classifiers in a sequence, reweighting the data instances each in turn to reflect the difficulty for each of them (Biamonte, et al., 2017).

Stacking is a more complex EL method that involves training multiple models and then combining their predictions using a meta-model. It is complicated where it involves training several base models on the same dataset and having a supervisor (meta) model that learns how to combine the best predictions of the base models (Gudivada, 2016). Examples of stacking algorithms include Meta-Decision Tree (MDT), which uses a decision tree as the meta-model, and Stacked Generalization (SG), which uses a linear regression model as the meta-model. However, overfitting is common in stacking where the base model is too complex.

Generalized Linear Models (GLMs) is a meta-learner for stacking EL, it is a type of model that can be apply in various kinds of outcomes, such as regression and classification outcome (Peterson, Baker, & McGaw, 2010). GLMs combine the predictors from multiple models into a single model, whereas other EL methods combine the outputs of multiple models. GLMs are often preferred for predicting clinical outcomes due to the major weakness of ensemble predictors, which typically produce "black box" predictions that are difficult to interpret in terms of underlying features.

GLM is a type of forward-selected regression model that results in highly interpretable predictors (Song, et al., 2013). (Kwon, et al., 2019) applied the stacking EL technique for classifying breast cancer based on Korean women cohort, where the study proves that gradient boosting model and GLM as a meta-learner shows better performance than single classifiers. A recent study by Kasim S, et al. (2023) focused on predicting in-hospital mortality in Asian women post-STEMI using ML and stacked EL using the same NCVD dataset and the models were compared to the conventional TIMI risk score, proven that ML and EL techniques provided more accurate classifications for Asian women with STEMI than traditional methods.

In this study, stacked EL approach was applied. The models mentioned earlier served as our base learners, which were then integrated and stacked together. Subsequently, the GLM was utilized as the meta learner to enhance the predictive capability of the combined model.

2.8.1.8 Summary of Machine Learning Algorithms

Table 2.12 below provides a summary of the ML algorithms applied in this study, highlighting both their strengths and drawbacks.

Table 2.12: Summary of ML algorithms that applied in this study.

Machine Learning Algorithms	Developed By	Strengths	Drawbacks
Linear Regression	Francis Galton, 1886	- Simple and easy to understand. - Computationally inexpensive to train. - Scales well to large datasets. - Coefficients of the model can give insight about the importance of features. - Suitable for continuous output (regression problems)	- Assumes linear relationship between the features and the target. - Sensitive to outliers. - May suffer from overfitting or underfitting. - Cannot model complex relationships without transformation.

Table 2.12, continued.

Machine Learning Algorithms	Developed By	Strengths	Drawbacks
Logistic Regression	David Cox, 1958	- Simple and interpretable model. - Fast to train and predict. - Scales well to large datasets. - Predictions can be made quickly. - Widely used and well-studied. - Many resources and tools available for it.	- Linear model, limited to linear relationships between the features and the target. - Cannot capture more complex relationships between the features and the target. - Sensitive to the scale of the features. - Sensitive to the presence of outliers, it can have a large influence on the model. - Only for classification model.
Support Vector Machine (SVM)	Vladimir Vapnik and his colleagues, mid-1990s	- Effective in high-dimensional spaces - Effective when the number of features is greater than the number of samples. - Effective where the data is heavily imbalanced. - Versatile, and different kernel functions can be used to specify the similarity between samples. - Used in a variety of applications.	- Sensitive to the choice of kernel and parameters. - Difficult to find the right combination. - Slow to train and predict, particularly on large datasets. - Do not provide probabilities for the outcomes, which can be useful in certain contexts.

Table 2.12, continued

Machine Learning Algorithms	Developed By	Strengths	Drawbacks
Random Forest (RF)	Leo Breiman and Adele Cutler, early 2000s	- Effective models. - Able accurately predict outcomes in many cases. - Fast to train and predict. - Handle large datasets with many features. - Handle data with missing values and data with imbalanced classes. - Provide a good balance between bias and variance. - Generally, have good generalization performance. - Easy to use. - Require little pre-processing of the data.	- Difficult to interpret. - Hard to understand why a particular prediction was made. - Difficult to extract insights from the model. - Prone to overfitting if not properly tuned. - Slower to train and predict especially when working with very large datasets.
Extreme Gradient Boosting (XGBoost)	Tianqi Chen and his colleagues, 2014	- Fast and efficient. - Able to handle large-scale data sets. - Can effectively deal with high-dimensional data. - Has a number of useful hyperparameters that can be tuned to improve model performance.	- Complex algorithm. - Difficult to interpret. - Prone to overfitting. - Sensitive to hyperparameters.

Table 2.12, continued

Machine Learning Algorithms	Developed By	Strengths	Drawbacks
Naïve Bayes (NB)	Thomas Bayes	- Simple and efficient - Performs well on high dimensional datasets. - Small amount of data is required for training data. - Handles irrelevant features well. - Works well in classification problems.	- Assume independence between features. - Does not work well with numerical values. - Sensitive to outliers. - Requires careful pre-processing and feature selection to ensure optimal performance.
Ensemble Learning (EL)	Various	- Improve the performance of the individual models by combining them. - Robust and less prone to overfitting. - Able to handle wide range of data types and structures. - Stable and less noisy.	- Computationally expensive, especially large dataset. - Difficult to interpret and understand the underlying relationship data. - Requires careful selection and tuning of individual models to ensure optimal performance.

2.8.2 Explainable Artificial Intelligence (XAI)

Explainable Artificial Intelligence (XAI) is an emerging research field that aims to bring explanation to ambiguous and extremely complex ML models (Weber, et al., 2022). The interpretation of ML models is a crucial aspect of constructing predictive models in the field of data science. While ML models excel at discovering complex patterns in large datasets, they frequently suffer from the "black box" problem, in which they are difficult to interpret and offer little insight into the underlying relationships between the input features and the output predictions (Rodríguez-Pérez & Bajorath, 2019). Ekanayake et al. (2022) investigated the complexities of machine learning's 'black box' nature, focusing on predicting compressive strength with SHAP and XGBoost. Their investigation revealed how these forecasts capture the intricate relationships between the constituents. SHAP also provides unified measures on feature importance and the impact of a variable for a prediction.

To improve the transparency and accountability of ML models, model interpretation techniques have been developed that aim to provide more intuitive explanations for model predictions. Interpretation methods are often divided into two categories: global and local interpretation. Global interpretation methods aim to provide an overall understanding of the model by analyzing its overall performance and feature importance.

2.8.2.1 Shapley Additive exPlanations (SHAP)

Shapley Additive exPlanations or SHAP explainer is a mathematical method, that uses Shapley values to explain how individual predictions are made by the ML model. SHAP was introduce and developed by (Lundberg & Lee, 2017). SHAP provides a global interpretation of a model by assigning an important score to each feature based on its contribution to the model's output. It considers all possible feature combinations and their impact on the model's output. SHAP addresses the 'black box' issue in machine learning, as depicted in Figure 2.11

The SHAP values for a particular feature indicate how much that feature contributed to the predicted outcome, either positively or negatively. Positive SHAP values indicate that the feature increased the prediction, while negative SHAP values indicate that the feature decreased the prediction. The sum of the SHAP values for all features adds up to the difference between the actual predicted value and the average predicted value across all samples in the dataset.

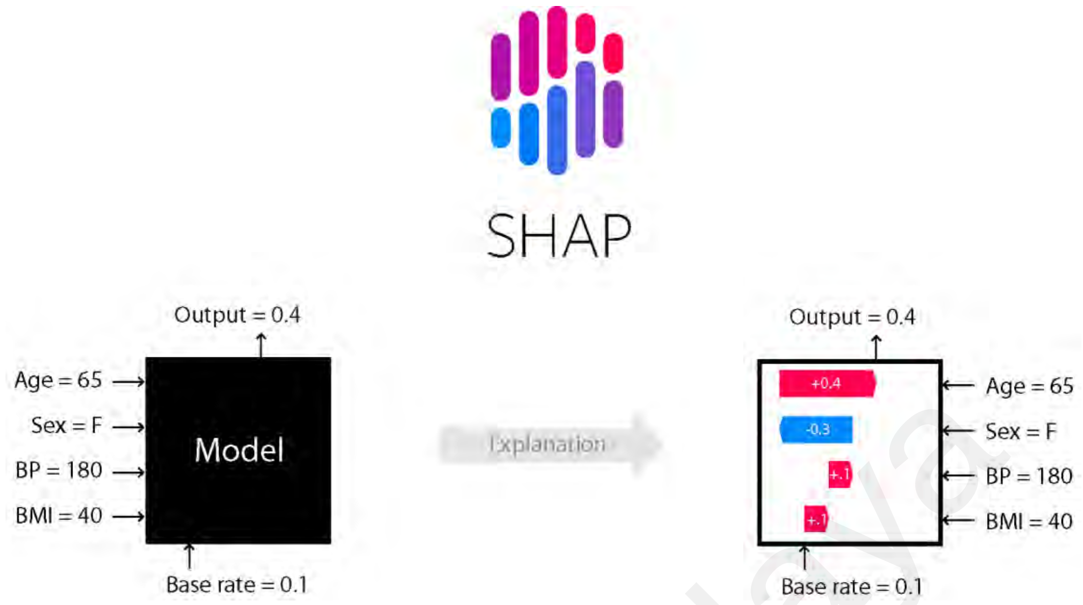


Figure 2.11: Illustration of SHAP interpreting the ML "black box" nature (Photo sourced from SHAP, 2017).

The SHAP summary plot is a visualization tool that helps to interpret the SHAP values. It shows the contribution of each feature to the final prediction in a single plot. The horizontal axis represents the SHAP value, and the vertical axis represents the feature value. The points in the graph represent individual data points, and their position on the vertical axis indicates the value of the feature for that data point. The color of the points indicates the value of another feature that is correlated with the feature being plotted. Red indicates high feature value, whereas blue represents low feature value. Figure 2.12 below illustrates an example of the SHAP graph adapted from (SHAP, 2017).

Thus, XAI is addressed by SHAP to provide easily understood graphical interpretations of results obtained from conventional AI approaches.

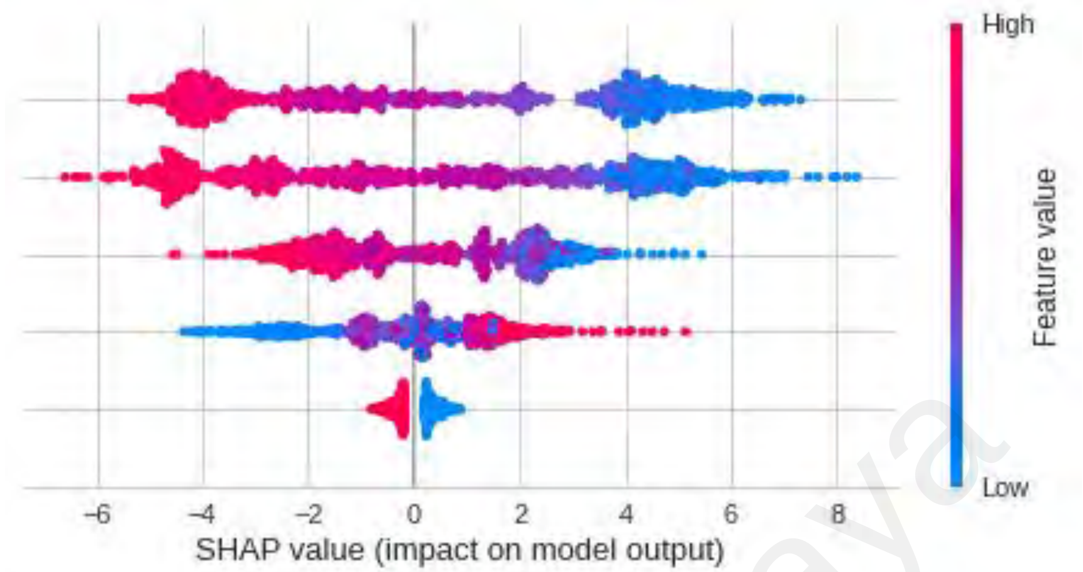


Figure 2.12: Illustrations of SHAP summary plots (Photo sourced from Tran K., 2021).

SHAP has grown in popularity and has become a widely used method for interpreting the ML models in various field (Lin & Gao, 2022), prevention of hypoxemia during surgery, where impact of each feature on the model is represented using Shapley values (Lundberg, et al., 2018), spatial drought prediction model interpretation that uses SHAP plots to examine predictor interactions for various drought conditions (Dikshit & Pradhan, 2021). In Liu, et al. (2022) study about diagnosis of Parkinson's disease, to address the problem of high feature dimensionality of Parkinson's disease in medical data, SHAP value is apply for feature selection of Parkinson's disease. Similarly, Kuno, et al. (2022) uses the SHAP method to identify and interpret the important variables that are associated with in-hospital mortality for COVID-19 patients.

2.8.3 Machine Learning (ML) Performance Evaluation

ML performance evaluation is a critical step in the model development process. Its performance is evaluated to determine its effectiveness and identify areas for improvement, it also helps to select the most appropriate model for further development. According to Kim et al. (2017), the commonly used metrics for evaluating the performance of a diagnostic model included confusion matrix, accuracy, sensitivity, specificity, precision, F1 Score, recall, and area under the ROC curve (AUROC).

However, the performance evaluation for regression and classification problems may differ. In regression problems, the most commonly used performance evaluation metrics include mean absolute error (MAE), and root mean squared error (RMSE), while in classification problems, accuracy, precision, and AUROC are commonly used. The following metrics are used in this study to evaluate model performances (Sun, et al., 2021).

2.8.3.1 Regression Performance Evaluation

(a) Mean Absolute Error (MAE)

Mean absolute error (MAE) is used to calculate the accuracy of continuous variables generated from regression model (Zhou et al., 2019). MAE represents the average magnitude of the error in a set of predictions (Nadakinamani, et al., 2022), MAE represents the average absolute difference between the predicted and actual values. The smaller the MAE, the better the performance of the model. MAE calculated using the equation as shown below,

$$MAE = \frac{1}{n}\sum_{J=1}^{n}|y_i - \widehat{y_i}| \qquad (2, 5)$$

Where $\widehat{y_i}$ are the predicted values, y_i are the observed values and n is the number of observations.

(b) Root Mean Square Error (RMSE)

The Root Mean Square Error (RMSE) is one of the performance indicators for ML regression model, it measures the average difference between values predicted by a model and the actual values. It provides an estimation of how well the model can predict the target value. Thus, the RMSE value ranges from 0 to infinity, where the value is closer to 0 the better the performing the regression model (Peter J., et al., 2022). RMSE is the square root of the mean square error between the predicted and actual values, it is calculated as shown in the equation below:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\widehat{y_i} - y_i)^2}{n}} \tag{2,6}$$

Where $\widehat{y_i}$ are the predicted values, y_i are the observed values and n is the number of observations.

In recent years, RMSE has been widely used in the medical domain for various applications such as disease prediction, diagnosis, and prognosis. For instance, a study by Nadakinamani, et al. (2022) used MAE and RMSE to evaluate the performance of a ML model for predicting the risk of CVD, the random tree shows the best result of 0.0011 and 0.0231 respectively, indicating good predictive performance.

2.8.3.2 Classification Performance Evaluation

(a) Confusion Matrix (CM)

The confusion matrix (CM) was utilized for the performance evaluations. CM the number of actual and predicted values, it can be applied to binary classification as well as for multiclass classification problems. CM is a widely used evaluation metric in ML, particularly in the medical domain (Demir, 2022; Asif, et al., 2021; Hossen, et al., 2021; Imamovic, Babovic, & Bijedic, 2020).

CM consists of four basic characteristics (numbers) that are used to define the measurement metrics of the classifier (Singh, et al,, 2021). These four numbers are:

1. TP (True Positive): TP represents the number of patients who have been properly classified to have malignant nodes, meaning they have the disease.
2. TN (True Negative): TN represents the number of correctly classified patients who are healthy.
3. FP (False Positive): FP represents the number of patients who have been misclassified as having the disease but are actually healthy. FP are also known as *Type I errors*.
4. FN (False Negative): FN represents the number of patients misclassified as healthy but actually they are suffering from the disease. FN is also known as *Type II error*.

The output True Negative (TN) indicates that the number of negative examples that were accurately classified as negative. Similarly, True Positive (TP) represents the number of positive examples that have been correctly classified. The term False Positive (FP) value, which is the number of actual negative examples misclassified as positive, whereas False Negative (FN) value, which is the number of actual positive examples misclassified as negative. (Kulkarni, et al, 2020)**.** CM also known as the error matrix, is depicted by a matrix describing the performance of a classification model on a set of test data as shown in figure 2.13 below (Sharma, et al., 2022).

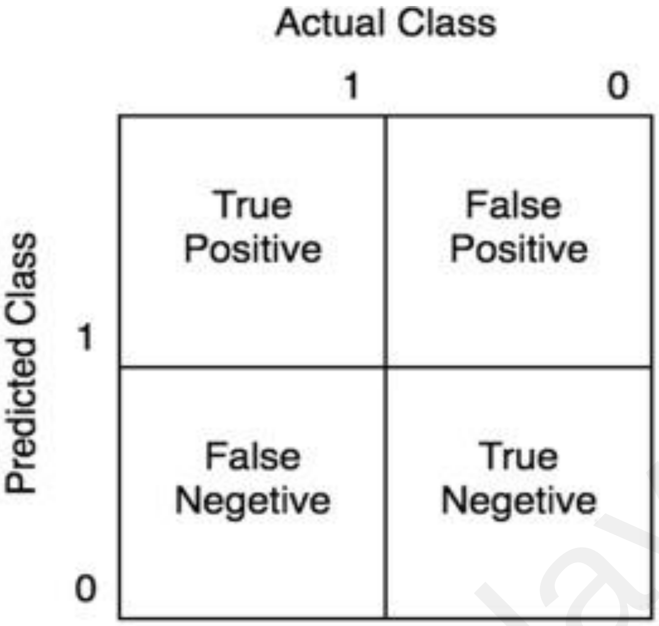


Figure 2.13: Confusion Matrix Plot (Photo sourced from Sharma, et al., 2022).

Performance metrics of an algorithm are accuracy, precision, sensitivity, specificity, and F1 score, which are calculated based on the above-stated TP, TN, FP, and FN.

(b) Accuracy (ACC)

One of the most commonly used metrics while performing classification is accuracy (Kulkarni, et al., 2020; Saura 2021). The accuracy is calculated by the number of correctly predicted from the total number of samples in the dataset. The accuracy of a model (through a confusion matrix) is calculated using the given formula below.

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN} \tag{2,7}$$

The result indicates the ratio of the sample to be correctly classified. Higher accuracy leads to a better performance in the model (Story & Congalton, 1986). However, sometimes accuracy can be misleading if used with imbalanced datasets, and therefore there are other metrics based on confusion matrix which can be useful for evaluating performance (Kulkarni, et al., 2020).

(c) Precision (P)

Precision of an algorithm is represented as the ratio of correctly classified patients with the disease (TP) to the total patients predicted to have the disease (TP+FP) (Singh, et al., 2021). In other words, it is the proportion of positive values that were correctly defined (Arjaria, et al., 2021). The formula of calculate precision is as follows:

$$Precision = \frac{TP}{FP + TP} \qquad (2, 8)$$

(d) Sensitivity (Sn) / Recall / True Positive Rate

Sensitivity is defined as the ratio of the predicted genuine positive cases to all positive cases and known as recall or true positive rate. Sensitivity is the ability of a test to identify those with the disease correctly. If the test has a high sensitivity, and the test result is negative, it is nearly certain that they do not have the disease (Jain & Singh, 2018).

$$Sensitivity = \frac{TP}{FN + TP} \qquad (2, 9)$$

(e) Specificity / True Negative Rate

Jain & Singh (2018) defined specificity as the ratio of true negatives to the sum of true negatives and false positives. Specificity is used to identify misclassifications in negative cases (Veropoulos, et al., 1999). The specificity formula is as follows:

$$Specificity = \frac{TN}{FP + TN} \qquad (2, 10)$$

(f) F1 Score

F1 score is also known as the F Measure. The F1 score states the equilibrium between the precision and the recall. The F-Measure presents a method for combining precision and recall into a single measure that can capture both of these features (Taha & Hanbury, 2015).

$$F1\,Score = \frac{precision\ \times sensitivity}{precision + sensitivity} \times 2 \qquad (2, 11)$$

(g) *Receiver Operating Characteristics (ROC) Curve*

The receiver operating characteristic (ROC) curve is the presentation plot derived from the confusion matrix, specificity, and sensitivity with the True positive rate (TP rate) versus the False positive rate (FP rate). The area under the ROC curve (AUROC) is an additional metric used to evaluate the algorithm's performance efficacy (Dutta, et al., 2023). AUROC 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). Figure 2.14 depicts the ROC for each of the three ML algorithms RF, XGBoost, and Decision Tree. The graph reveals that, among all these ML classifier algorithms, RF's AUROC was the greatest.

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). Ling, et al. (2003) conducted a study which has shown that AUROC is much more suitable than using accuracy for balanced and imbalanced data sets. In the study by Wallert, et al. (2017), AUROC is taken as the performance metric for their models developed as classes were heavily unbalanced, as it is not imbalance sensitive. With the values ranging from 0 to 1, 0.5 corresponds to random guessing, and any feature or variable with AUROC > 0.7 might be a potentially useful clinical classifier. However, the judgement is also made along with the consideration on base rate incidence, consequences of false negatives/positives, test risk, cost, etc. According to Seliya, et al. (2009), a large area under the curve is much preferable than a small area under the curve for a classifier.

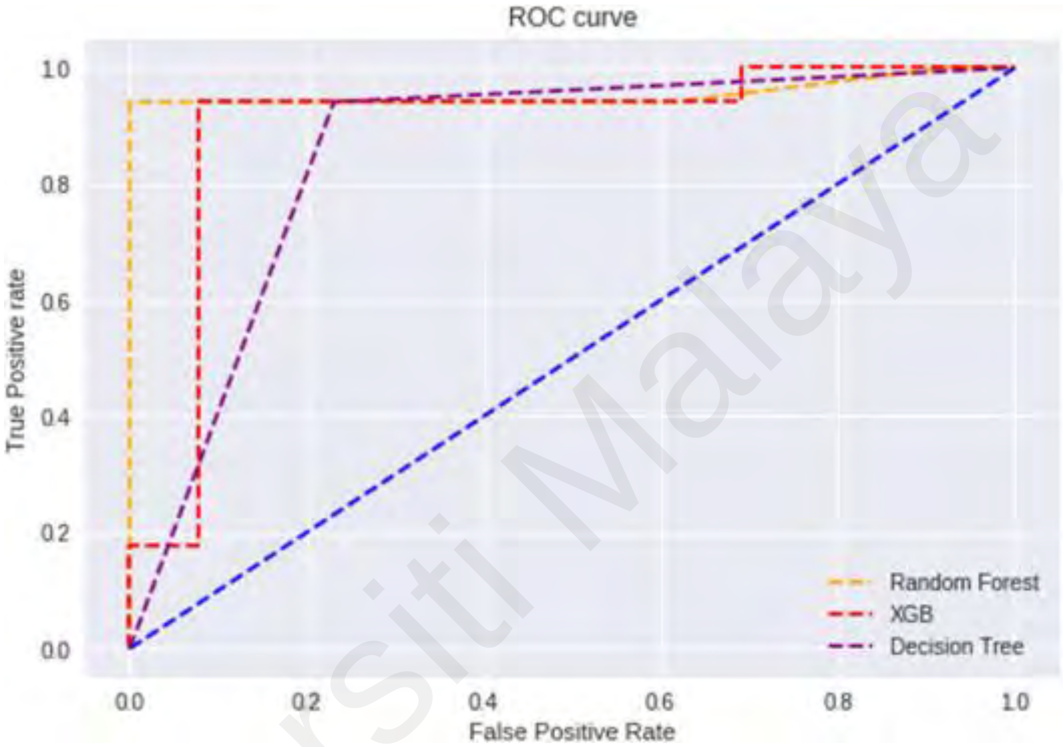


Figure 2.14: Example of ROC curve that shows the performance of the machine learning models (Photo sourced from Seliya, et al., 2009).

Table 2.13 below presents the comparison of model evaluation between regression and classification ML model.

Table 2.13: Comparison of model evaluation between regression and classification machine learning model.

Evaluation Metric	Description	Regression	Classification
Root Mean Square Error (RMSE)	The square root of the average of the squared differences between predictions and actual values.	✓	

Table 2.13, continued.

Evaluation Metric	Description	Regression	Classification
Mean Absolute Error (MAE)	The average of the absolute differences between predictions and actual values.	✓	
Area Under Curve (AUC)	The probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one.		✓
Sensitivity	Measures the proportion of actual positives that are correctly identified.		✓
Specificity	Measures the proportion of correctly predicted positive observations out of the total predicted positives.		✓
Accuracy	Measures the proportion of correct predictions over total predictions.		✓

(✓) indicates the metric is commonly used for the respective model type.

2.9 Existing Study on Machine Learning (ML) for CVD Patients in Presence of Air Pollution

The study of CVD remains intriguing. The literature shows that ML approaches have been widely used to predict mortality risk and risk of CVDs. However, despite the success of ML in CVD risk prediction, most existing studies do not take environmental factors, such as air pollution, into account. As a result, the mortality risk may be inaccurately estimated, leading to suboptimal clinical decision-making and potentially harmful outcomes for patients. This gap in the current research needs to be addressed in future studies.

However, with a well-developed model, the majority of the model does not further develop into a visualization system that fully utilizes the developed model; thus, in this study, the developed model is further developed and implemented into a web-based system that can generate predicted results for the user as an inference.

Table 2.14 below presents a summary of available studies on ML in CVD research, highlighting the effectiveness of ML models in predicting mortality risk and CVDs and able to produce promising result in compared to result produce by conventional statistical method and risk scoring method. However, no papers have been published on mortality risk in the presence of air pollution using the ML method, which is the identified as the research gap that aim to fill in this study.

Table 2.14: A summary of the available studies on machine learning in CVDs.

Research Journal and References	Data Sources	Number of Instances	Input Variables	Model Performances Metrics
Machine Learning Outperforms ACC/AHA CVD Risk Calculator in MESA. (Kakadiaris, et al., 2018)	MESA (the Multi-Ethnic Study of Atherosclerosis)	6459	11 variables	AUC Score: ACC/AHA (0.71) SVM (0.92)
Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants. (Alaa, et al., 2019)	UK Biobank	423,604	473 variables	AUC-ROC: Framingham Risk Score (0.724) Cox PH model (7 core-Var :0.734) Cox PH model (all: 0.758) SVM (0.709) RF (0.730) Neural networks (0.755) AdaBoost (0.759) Gradient Boosting (0.769) Auto-Prognosis (7 core-Var: 0.744) (369 non-Lab Var: 0.761) (104 Lab Var: 0.735) (All: 0.774)

Table 2.14, continued.

Research Journal and References	Data Sources	Number of Instances	Input Variables	Model Performances Metrics
Machine learning adds to clinical and CAC assessments in predicting 10-year CHD and CVD deaths. (Nakanishi, et al., 2021)	Coronary Artery Calcium (CAC) Consortium	66,636	77 Variables	**AUC** ML (0.845) ASCVD Risk Score (0.821) CAC Risk Score (0.781) ML CT (0.804) ML CHD death (0.860) ASCVD death (0.835) CAC death (0.816) ML CT (0.827)
Multiclass machine learning vs. conventional calculators for stroke/CVD risk assessment using carotid plaque predictors with coronary angiography scores as gold standard: a 500 participants' study. (Jamthikar, et al., 2021)	Kingston General Hospital's Cardiac Catheterization Laboratory, Ontario, Canada	500	39 variables	**AUC** FRS (0.62) SCORE (0.60) ASCVD (0.59) SVM-rbf (0.92) RF (0.94) XGBoost (0.93)

Table 2.14, continued.

Research Journal and References	Data Sources	Number of Instances	Input Variables	Model Performances Metrics
A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. (Dinh, et al., 2019)	National Health and Nutrition Examination Survey (NHANES) dataset	5000	131 variables	**AUC:** RF (0.829) XGBoost (0.830) SVM (0.816) LR (0.822) Ensemble (0.831)
Identification of cardiovascular diseases using machine learning (Louridi, Amar, & El Ouahidi, 2019)	UCI heart disease dataset	303	13 variables	**Accuracy for 70:30 Train-Test split** SVM (linear: 81.01%) (RBF: 61.11%) KNN (84.44%) Bayes Naif (83.33%)

Table 2.14, continued

Research Journal and References	Data Sources	Number of Instances	Input Variables	Model Performances Metrics
In-hospital risk stratification algorithm of Asian elderly patients (Kasim, et al., 2022)	Malaysian National Cardiovascular Disease Acute Coronary Syndrome (NCVD-ACS) registry	3991 elderly patients	50 variables	**AUC Score** TIMI (0.75) LR (0.91) RF (0.91) SVM (0.91) XGB (0.89) DL (RF selected variables: 0.956)
In-hospital mortality risk stratification of Asian ACS patients with artificial intelligence algorithm. (Kasim, et al., 2022)	Malaysian National Cardiovascular Disease Acute Coronary Syndrome (NCVD-ACS) registry	68528 patients	54 variables	**AUC Score - STEMI** DL (SVM selected var: 0.96) DL (RF selected var: 0.96) NSTEMI Patient DL (SVM selected var: 0.96) DL (RF selected var: 0.95)

Table 2.14, continued.

Research Journal and References	Data Sources	Number of Instances	Input Variables	Model Performances Metrics
Deep-learning-based risk stratification for mortality of patients with acute myocardial infarction (Kwon, et al., 2019)	Korean Myocardial Infraction (KorMI) registry	22874	14 variables	**AUROC - STEMI patient** DL (0.905) RF (0.890) LR (0.873) GRACE (0.851) ACTION (0.852) TIMI (0.781) **AUROC - NSTEMI patient** DL (0.870) RF (0.851) LR (0.845) GRACE (0.810) ACTION (0.806) TIMI (0.593)
Acute coronary syndrome prediction in emergency care: A machine learning approach (Emakhu, et al., 2022)	Urban emergency department (ED) state of Michigan	362138	58 variables	**AUC Result** XGBoost (0.97) Gradient Boosting (0.98) Adaboost (0.85)

Table 2.14, continued.

Research Journal and References	Data Sources	Number of Instances	Input Variables	Model Performances Metrics
A Stacking Ensemble Prediction Model for the Occurrences of Major Adverse Cardiovascular Events in Patients with Acute Coronary Syndrome on Imbalanced Data (Zheng, et al., 2021)	Korea Acute Myocardial Infarction Registry National Institutes of Health (KAMIR-NIH)	13104	60 variables	**AUC Result** LR (0.6647) SVM (0.6585) KNN (0.8245) DT (0.9367) RF (0.9696) XGBoost (0.9696) AdaBoost (0.9569) Ensemble Stacking (0.9863)
Predicting Acute Onset of Heart Failure Complicating Acute Coronary Syndrome: An Explainable Machine Learning Approach (Ren, et al., 2022)	Guangdong Second Provincial General Hospital	1563	128 variables	**AUC Result** BRF (full features) (0.760) BRF (first 20% features) (0.785) LLR (full features) (0.604) LLR (first 20% features) (0.658)

Table 2.14, continued.

Research Journal and References	Data Sources	Number of Instances	Input Variables	Model Performances Metrics
Machine learning-based in-hospital mortality prediction models for patients with acute coronary syndrome (Ke, et al., 2022)	Emergency department of Fujian Provincial Hospital	6482	25 variables	**AUC Result** LR (0.884) GBDT (0.918) RF (0.913) SVM (0.896)
Predicting 30-Day Mortality after an Acute Coronary Syndrome (ACS) using Machine Learning Methods for Feature Selection, Classification and Visualisation (Aziida, et al., 2021)	Malaysian National Cardiovascular Disease Acute Coronary Syndrome (NCVD-ACS) registry	302 patients	54 variables	**AUC Score** RF-varImp-SBS (0.80) TIMI (0.60)
Mortality prediction of elderly Asian patients with acute coronary syndrome (ACS) using interpretable machine learning algorithm (Kasim S. S., et al., 2022)	National Cardiovascular Disease Database (NCVD)	4305	10 variables	**AUC Result** XGBoost-STEMI (0.8116) XGBoost-NSTEMI (0.8471)

2.10 Geospatial Mapping

Both statistical and ML models do not include visualization elements for better dissemination of information. Geographic Information Systems (GIS) is an essential scientific tool for health data processing, analysis of geographical distribution and variation of diseases, mapping, monitoring and management of health epidemics. Epidemiologic studies have adopted GIS to explore the health impact of air pollutants on asthma. GIS can significantly improve epidemiologic research in terms of definition of source and routes of possible exposure and approximation of environmental levels of target contaminants in the exposure assessment process (Nuckols, et al., 2004).

Data visualization is an essential tool for effective communication and interpretation of information. Visualization of information in terms of image is easier as compared to numerical values. Furthermore, vast availability of data requires effective and efficient ways to access and communicate information (de Vries, Land-Zandstra, & Smeets, 2019). Visualization is essential for data analysis and data representation (Xu, et al., 2010).

Google Earth provides a platform for a which supports a visualization system online from GIS. According to Silberbauer (2009), Google Earth interface allows rapid changes of scale from global to local and back and it is user friendly. Furthermore, Google's massive centralized spatial database keeps updates for users.

According to McGranaghan (1993), graphic visualization is essential in setting communication objectives because different data users will have different data quality visualization needs. Map users rely on graphic quality to assess map accuracy and data quality.

The goal of visualization in this study is to combine the strengths of human vision, creativity, and general knowledge with the storage capacity and the computational power of modern computers to explore extensive environmental and healthcare data. This is implemented in this study by presenting graphical representations of the data to the user, which allow users to interact with the data to gain insight and to draw conclusions quickly (Keim, 2005).

Graphical analysis is necessary to identify patterns, trends, and other features that are not noticeable from numerical summaries. Data visualization is important because it allows for the clear and effective communication of information through graphical means. It is useful to analyze data and identify geographic areas with the highest pollution rates and the most number of patients.

The use of geospatial analysis in Parkinson's disease (PD) research has exhibited 6% of spatial dependence for a small but significant proportion of patients in the Canton of Geneva, Switzerland, provides new insights into environmental epidemiology research in PD (Fleury, et al., 2021). In addition, geospatial analysis is essential for identifying potential environmental risk factors and developing effective public health interventions, as demonstrated in Deaths from cholera in Soho, London (late July to end of September 1854) which revealed the epidemiology of the disease and demonstrated the application of geospatial analysis by highlighting the shortest path principle followed by local residents when they drank water from a contaminated pump (Walford, 2020).

In the case of COVID-19, there is substantial spatial variation in the spread of the disease and localized distinctions in transmission. Infections propagate more rapidly in urban and well-connected regions than in rural and poorly connected regions, thus, to design effective control programs, it is essential to comprehend the local geospatial variation of COVID-19 transmission (Cuadros, et al., 2020). Overall, geospatial mapping has become a valuable tool for disease studies, allowing for better understanding and control of disease transmission patterns.

Hence, in this study, the impact of air pollution on ACS patients was studies using ML approach, comparing the ML model performance with existing risk scoring method. Moreover, GIS concept in combination with Google map will also be used to provide graphical outcome of air pollution and ACS in Malaysia.

2.11 System Development Life Cycle (SDLC)

The System Development Life Cycle (SDLC) referred as the systematic approach for plan, analysis and designing an information system (Kendall & Kendall, 2002). It is a systematic approach to software development that guides the development team through the planning, deployment, and maintenance phases of the software development process. Each of the six phases of the SDLC has its own set of activities and deliverables. These phases include Feasibility study, System investigation, System analysis, System design, System implementation, Review, and system maintenance (Stefanou, 2003). Figure 2.15 illustrates the SDLC workflow.

(a) Feasibility Study

The feasibility study examines existing systems in consideration of emerging demands and considers alternative solutions. System analysts determine whether a newly identified system or application requirement is essential for the organization (Stefanou, 2003).

(b) System Investigation

System investigation phase is a comprehensive examination of the functional requirements, performance, and limitations of the existing system, if one is presently in operation, and the intended system's requirements. At this juncture, a more comprehensive analysis will be conducted, taking into consideration the data types, volumes, and transactions that the new system will need to process.

(c) System Analysis

In this phase of SDLC and this phase aims to understand the client's requirements of the system/application. It is used to identify what is needed for the system (Radack, 2009). All the information needed to be processed, transmitted, and stored is evaluated and the purpose of the system is documented in this phase (Conrad, Misenar, & Feldman, 2016).

(d) System Design

In this phase, alternative technical solutions are evaluated, and the hardware, software, human resources, and procedures for the new system are specified. This phase also includes creating a prototype of the system and reviewing it with stakeholders to ensure it meets their needs.

(e) System Implementation

System implementation phase is the activity of installing according to specifications and delivering into operation a computer system, it involves programming, system testing, documenting, and delivering the system into operational use.

During this phase where coding, debugging, and developing the software occurs in this phase. This includes creating the user interface (UI) and integrating the system with other systems as required (TutorialsPoints, 2023).

Testing is a crucial subphase for determining the quality and efficacy of the developed information system. A few categories of tests include unit test, integration test, volume test, system-test and user-acceptance test.

(f) Review and System Maintenance

Once the system has been implemented, it is reviewed to ensure it meets the specified requirements. This includes testing the system and bugs, performing user acceptance testing, and conducting a final review with stakeholders.

The post-implementation evaluation and review activities are essential for determining the system's usefulness, any necessary modifications, and the extent to which it meets the project's objectives.

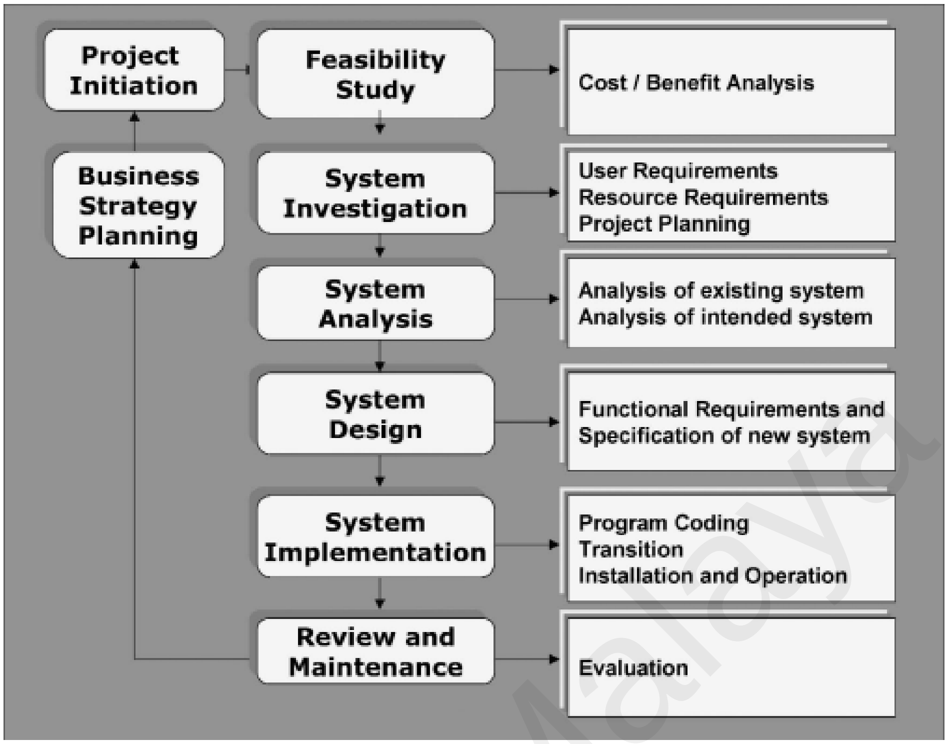


Figure 2.15: The conventional SDLC model (Photo sourced from Stefanou, 2003).

2.11.1 SDLC Methodology

Software Development Life Cycle (SDLC) methodologies are structured approaches to developing software systems. There are several types of SDLC methodologies, including prototyping, waterfall, Rapid Application Development (RAD), Dynamic Systems Development Method (DSDM). Each methodology has its unique set of characteristics and phases that dictate how software development should proceed from planning to deployment.

In this study, the prototyping methodology is utilized in developing the web system, which is further explained under the following subsection 2.11.1.1.

2.11.1.1 Prototyping

Prototyping is a software development methodology that emphasizes the use of continuously refined working models based on end-user feedback. Most frequently, prototyping is used to develop systems with substantial end-user interaction and complex user interfaces (Naimish, 2023).

According to Earl (1978) study stated that prototype methodology is viewed as a "process-enabler" and a tool for action researchers to learn more about information system design. Since it requires the developers to create an initial prototype of the software or application before developing the final product, it involves constant discussion between the end-users or stakeholders, which this model can help to reduce the risk of developing of software product that does not meet the requirements. The phase of the prototyping model is explained as follows:

1. Requirement gathering and analysis:

   The objectives of the system are precisely defined. Discussion and interviews are conducted between the developers and system users to determine their requirements for the system.

2. Design

   Design is the second phase that comprises a preliminary design of the system. The system fundamental design is established, where it only provides the user with a quick overview of the system. This phase facilitates the development of the prototype.

3. Build a Prototype

   During this stage, the prototype is developed based on the quick design created in the previous stage. The prototype is to support and validate the knowledge gained during the design stage.

4. Initial user evaluation

   The developed prototype is presented to the end-user for preliminary testing. In addition, it permits the developers to evaluate the performance of the initial model, thereby identifying its strengths and weaknesses. End-users and developers interact at this stage, discussing the system prototype and providing feedback on the design and functionality. This is done to ensure that the final product fulfils the requirements and expectations of the end user.

5. Refining prototype

The process of refining a prototype is a crucial step in the software development process, which allows for the iterative improvement of the design based on user feedback and suggestions. This process involves making changes and improvements to the prototype based on user feedback, such as adjustments to the user interface and functionality. Once the prototype has been refined, it is presented to the user again for evaluation. If the user is satisfied with the upgraded model, a final system based on the approved final type is created.

6. Implement Product and Maintenance

After the end-users are satisfied with the refined prototype developed moving on to the development phase of the system. The final phase where the final system was fully tested and distributed to production after it was developed.

The prototyping methodology is selected because it allows developers to quickly create a working prototype of the software, which can be used to test and refine the system's requirements. This methodology also enables in identifying potential issues early in the development process, which can save time and money in the long run. Figure 2.16 below shows an illustration of the prototype model development cycle.

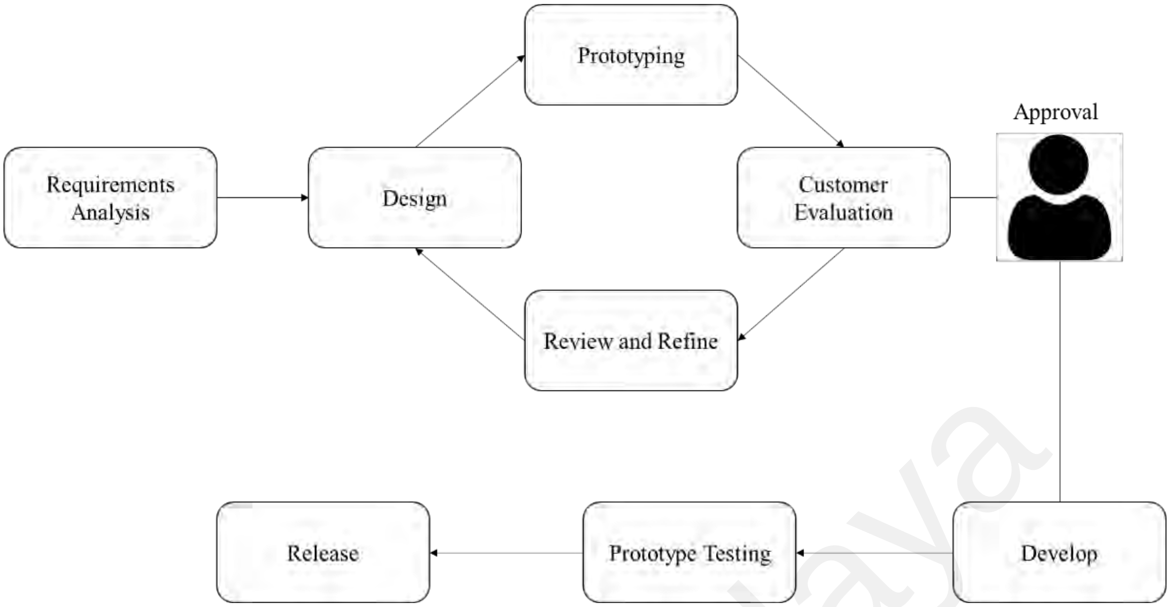


Figure 2.16: Prototype development cycle.

2.11.2 Comparison of SDLC Methodology

In section 2.11.2, we will compare the prototyping methodology with commonly used SDLC methodologies. The strengths and weaknesses of each methodology are evaluated and determined. This analysis will help us to make an informed decision about which SDLC methodology to use for our software development project (Alshamrani & Bahattab, 2015).

Table 2.15 below summarizes the advantages and disadvantages of software development methodologies and table 2.16 below shows the overview of existing SDLC methodology.

Table 2.15: Summary of advantages and disadvantages of existing software development methodologies.

Software Development Methodology	Advantages	Disadvantages
Waterfall Model	- Simple - Well-understood - Well-defined stage - Stable - Easy to manage. - Resources and expertise are available. - Easy for testing and analysis - Easy documentation - Suitable for small projects	- Rigid - High risk and uncertainty - Not suitable for object-oriented projects. - Not suitable for long and ongoing project. - Cannot allow modification. - The project should be precise. - One-time project
Agile Development Software Model	- Adaptive approach - Allow changes and modification. - Allow direct communication. - Able to fix bugs quickly. - Improve quality and fast review. - Fast delivery - User focused - Accept uncertainty	- Poor documentation - The outcome is not clear. - Get side-tracked easily. - Unable to complete within the allocated time. - Additional cost - Time-consuming

Table 2.15, continued

Software Development Methodology	Advantages	Disadvantages
Rapid Application Development (RAD)	- Focus on the all the advantages of development. - Fast delivery - Low Cost - High quality outcomes - Encourage feedback. - Improvement can be done	- Dependent on developer team - Modular system - Required skilled expertise. - Complex - Not suitable for small projects.
Dynamic System Development Model (DSDM)	- Iterative - Evolutionary - Incremental - Fast delivery of functionality - Easy access for users and developers - Results are direct and visible. - Users are actively involved.	- Costly - Not suitable for small projects

Table 2.15, continued.

Software Development Methodology	Advantages	Disadvantages
Prototype	- Fast delivery of working software - Enables ongoing feedback and collaboration between developers and users. - Direct and visible result	- Not suitable for small projects - Potential for scope creep if there is no clear understanding of requirements and goals. - Lead to technical dept if is not properly designed and tested.

Table 2.16: SDLC Methodologies Overview

SDLC Methodology	Waterfall	AGILE	RAD	DSDM	Prototype
Specification of All the Requirements in the beginning	Yes	Not all and Frequently Changed	Not all and Frequently Changed	Not all and Frequently Changed	Not all and Frequently Changed
Long term project	Inappropriate	Appropriate	Appropriate	Appropriate	Appropriate
Complex Project	Inappropriate	Appropriate	Appropriate	Appropriate	Appropriate
Frequently Changed Requirements	Inappropriate	Appropriate	Appropriate	Appropriate	Appropriate
Cost	Low	High	High	High	High
Cost estimation	Easy to estimate	Difficult	Difficult	Difficult	Difficult
Flexibility	Low	High	High	High	High
Simplicity	Simple	Moderate	Complex	Complex	Moderate

Table 2.16, continued.

SDLC Methodology	Waterfall	AGILE	RAD	DSDM	Prototype
Supporting High Risk Project	Inappropriate	Appropriate	Appropriate	Appropriate	Appropriate
Guarantee of Success	Less	High	Moderate	High	High
Customer Involvement	Low	High, after each iteration	High	High	High
Testing	Late	Execute during implementation phase	Fast	After each phase	After prototype is developed
Maintenance	Least maintainable	Maintainable	Least maintainable	Maintainable	Least maintainable
Ease of Implementation	Hard	Easy	Easy	Easy	Easy

2.12 Existing Web Application

Numerous digital systems have been developed to combat the growing incidence of CVD and the consequential impact it has on human health (Cornell, et al., 2023; Feigin, et al., 2022; Gjeka, et al., 2021; Urrea & Venegas, 2020; Brown, 2005). In addition, the negative impacts that air pollution has on human health (Brook R., et al., 2004; Kumar, et al., 2023) have further spurred the demand for geospatial analysis tools that can assist in identifying groups that are at risk and provide information that can inform targeted actions (Mazeli, et al., 2023; Zhalehdoost & Taleai, 2022).

In this section, the existing web systems are reviewed that are related to cardiovascular risk calculators and air pollution geospatial maps. Evaluating these existing systems is necessary to gain ideas and examples for developing cardiovascular web systems. Although these web-based cardiovascular risk tools were generally simple to use, only one-third provided risk modification advice. A well-chosen online cardiovascular risk assessment tool can help patients manage their health (Roshan, et al., 2023).

2.12.1 Existing Mortality Prediction Calculator Web Application

2.12.1.1 The Cleveland Heart Disease Risk Calculator

The Cleveland Heart Disease Risk Calculator is a tool developed by the Cleveland Clinic Department of Quantitative Health Sciences, United States. The department is a multidisciplinary group of biostatisticians, epidemiologists, outcomes researchers, database developers and programmers using biomedical research to improve patient care. Apart from heart disease risk calculator, the team also developed various risk calculators for other diseases such as brain cancer, bladder cancer, colorectal cancer, etc. The objective of the risk calculators developed is mainly to assist and convenience service only to physicians' medical advice.

The utility of the Cleveland Heart Disease Risk Calculator is significant, as it allows individuals to assess their risk of heart disease and make informed decisions about their health. Under the heart disease condition, Cleveland Clinic developed four risk calculators to meet the requirements and condition, which as follows:

(a) For Acute Coronary Syndrome Patients Recently Discharged

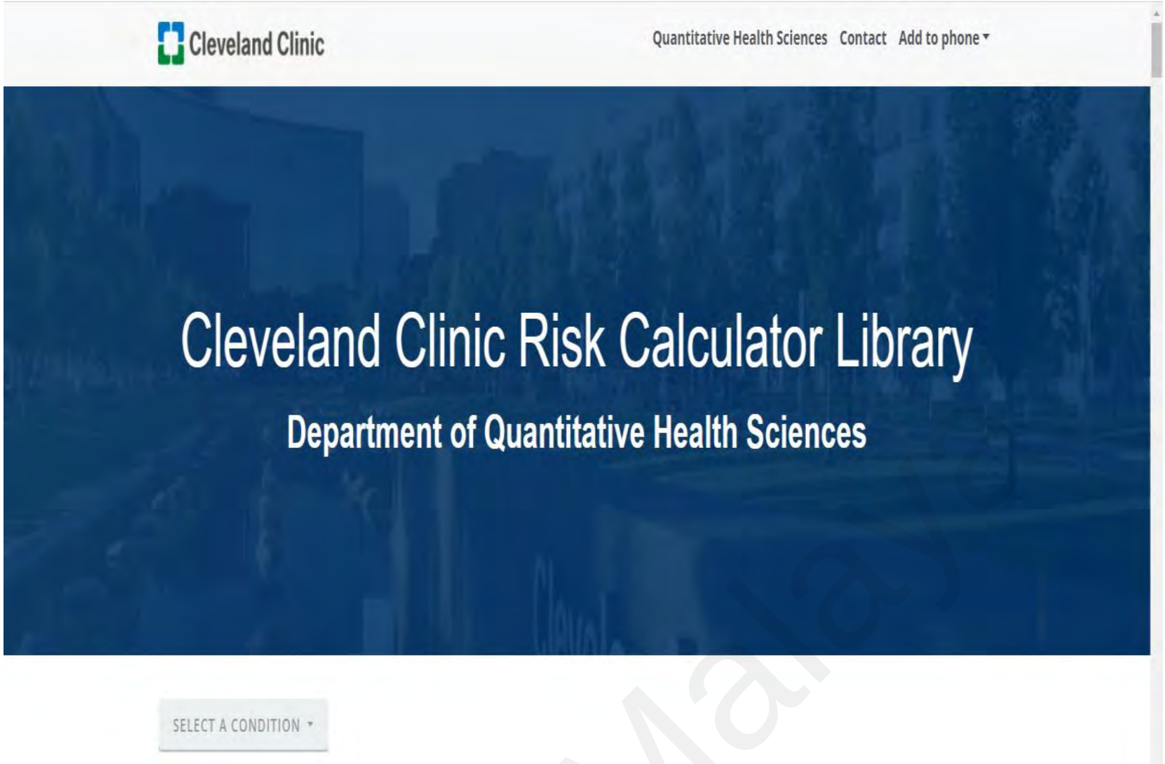
(b) For Patients Hospitalized with ACS Receiving PCI Treatment

(c) For Patients about to Undergo Transvenous Lead Extraction

(d) For Patient with Suspected Coronary Artery Disease and a Normal Electrocardiogram

The user is required to select a risk calculator according to their condition, we will review the risk calculator “For Acute Coronary Syndrome Patients Recently Discharged”, where it is relevant to our study. The 30 days risk prediction model was constructed using logistic regression and Cox proportional hazards were used to model the 1-year outcomes with 2681 patients. The c-indices for these models ranged from 0.73 to 0.82, then model developed is then incorporated into an easy-to-use online calculator (Kumbhani, et al., 2013). The calculator can be accessed online for free and the user interface is user-friendly, the results include 30 days risk of mortality, 30 days risk of myocardial infraction or revascularization, 1 year risk of mortality, 1 year risk of myocardial infraction or revascularization are displayed in a clear and understandable format, however the calculator only meant for patient with the age ranged from 30 to 85, and the metric unit could not be change. Figure 2.17 – Figure 2.20 are the screenshots of Cleveland Clinic ACS mortality risk calculator. The risk calculator could be access through this URL as follows:

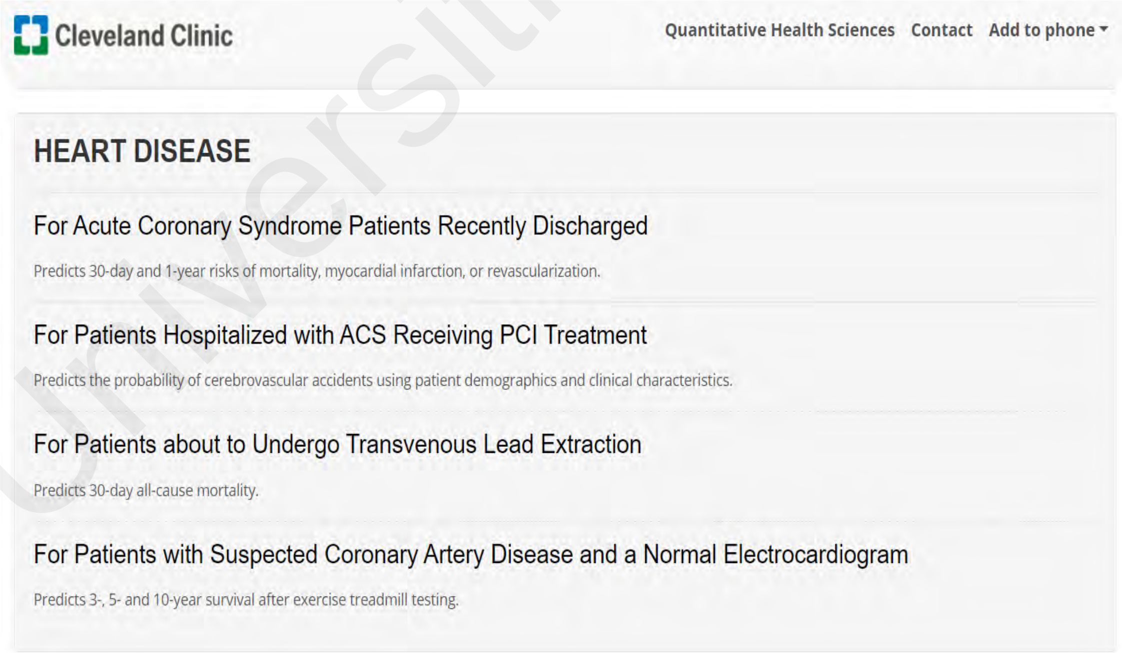
URL: https://riskcalc.org/.



Cleveland Clinic Risk Calculator Library

Department of Quantitative Health Sciences

Figure 2.17: The Cleveland Clinic Risk Calculator Library homepage.



HEART DISEASE

For Acute Coronary Syndrome Patients Recently Discharged

Predicts 30-day and 1-year risks of mortality, myocardial infarction, or revascularization.

For Patients Hospitalized with ACS Receiving PCI Treatment

Predicts the probability of cerebrovascular accidents using patient demographics and clinical characteristics.

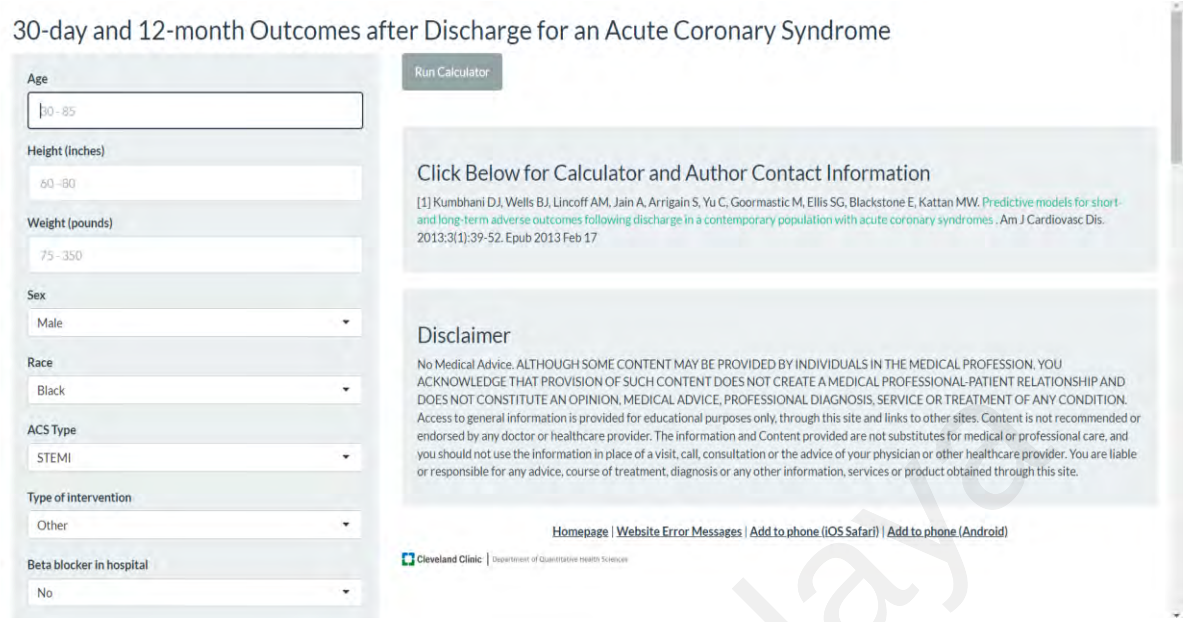
For Patients about to Undergo Transvenous Lead Extraction

Predicts 30-day all-cause mortality.

For Patients with Suspected Coronary Artery Disease and a Normal Electrocardiogram

Predicts 3-, 5- and 10-year survival after exercise treadmill testing.

Figure 2.18: Available heart disease calculator from Cleveland Clinic calculator.



30-day and 12-month Outcomes after Discharge for an Acute Coronary Syndrome

Age

30 - 85

Height (inches)

60 - 80

Weight (pounds)

75 - 350

Sex

Male

Race

Black

ACS Type

STEMI

Type of intervention

Other

Beta blocker in hospital

No

Click Below for Calculator and Author Contact Information

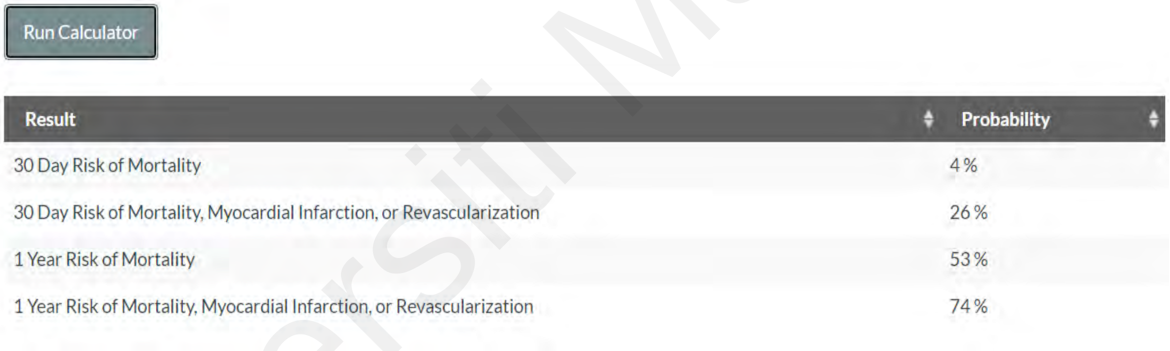
[1] Kumbhani DJ, Wells BJ, Lincoff AM, Jain A, Arrigain S, Yu C, Goormastic M, Ellis SG, Blackstone E, Kattan MW. Predictive models for short- and long-term adverse outcomes following discharge in a contemporary population with acute coronary syndromes . Am J Cardiovasc Dis. 2013;3(1):39-52. Epub 2013 Feb 17

Disclaimer

No Medical Advice. ALTHOUGH SOME CONTENT MAY BE PROVIDED BY INDIVIDUALS IN THE MEDICAL PROFESSION, YOU ACKNOWLEDGE THAT PROVISION OF SUCH CONTENT DOES NOT CREATE A MEDICAL PROFESSIONAL-PATIENT RELATIONSHIP AND DOES NOT CONSTITUTE AN OPINION, MEDICAL ADVICE, PROFESSIONAL DIAGNOSIS, SERVICE OR TREATMENT OF ANY CONDITION. Access to general information is provided for educational purposes only, through this site and links to other sites. Content is not recommended or endorsed by any doctor or healthcare provider. The information and Content provided are not substitutes for medical or professional care, and you should not use the information in place of a visit, call, consultation or the advice of your physician or other healthcare provider. You are liable or responsible for any advice, course of treatment, diagnosis or any other information, services or product obtained through this site.

Homepage | Website Error Messages | Add to phone (iOS Safari) | Add to phone (Android)
Cleveland Clinic | Department of Quantitative Health Sciences

Figure 2.19: Risk calculator for ACS patients recently discharged that predicts 3 days and 1-year risks of mortality, myocardial infraction, or revascularization.



Result	Probability
30 Day Risk of Mortality	4 %
30 Day Risk of Mortality, Myocardial Infarction, or Revascularization	26 %
1 Year Risk of Mortality	53 %
1 Year Risk of Mortality, Myocardial Infarction, or Revascularization	74 %

Figure 2.20: Result generated from the risk calculator that shows the probability of mortality, myocardial infraction, and revascularization.

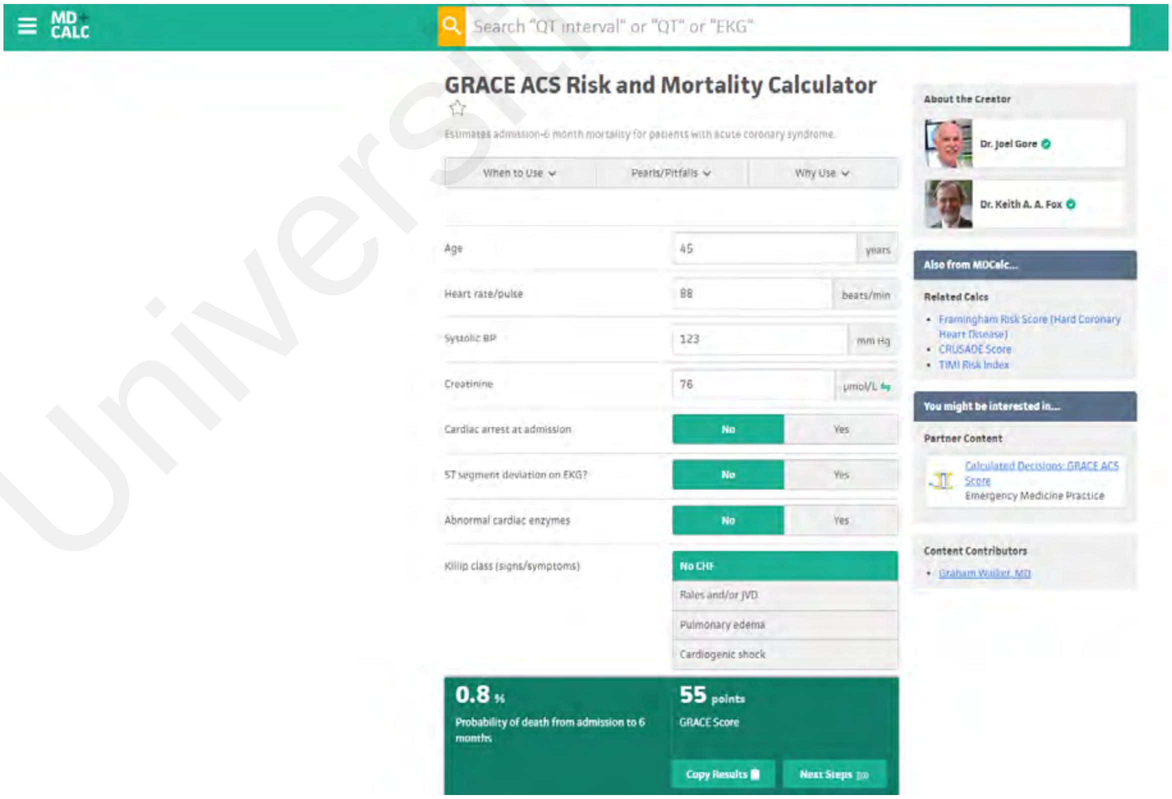
2.12.1.2 MDCalc for GRACE ACS Risk and Mortality Calculator

The GRACE ACS Risk and Mortality Calculator estimates the estimates the mortality rate from admission to six months for patients with ACS. This risk score calculator is created by Dr Joel Gore and Dr Keith A. A. Fox on MDCalc website (Gore & Fox, 2023).

The calculator is created based on (Fox, et al., 2006), the MDCalc for GRACE ACS Risk and Mortality Calculator consists of 8 indicators in the mortality prediction which includes the patient's age, heart rate, systolic blood pressure, creatinine level, cardiac arrest at admission, ST segment deviation, abnormal cardiac enzymes, and Killip class. GRACE risk scores come with some nonspecific features, such as patient's history, electrocardiogram, and troponin, which can be more objectively risk stratified and comes with better management and prognostication.

The website is simple and well-designed, the user just required to provide the input information, and the probability of death from admission to 6 months and the GRACE score points result will be calculated automatically as shown in Figure 2.21. Under the "evidence" tab, it will display the facts and figures interpreting the result generated as shown in Figure 2.22.

URL: https://www.mdcalc.com/calc/1099/grace-acs-risk-mortality-calculator



GRACE ACS Risk and Mortality Calculator

Estimates admission-6 month mortality for patients with acute coronary syndrome.

Age: 45 years

Heart rate/pulse: 88 beats/min

Systolic BP: 123 mm Hg

Creatinine: 76 µmol/L

Cardiac arrest at admission: No

ST segment deviation on EKG?: No

Abnormal cardiac enzymes: No

Killip class (signs/symptoms): No CHF

- Rales and/or JVD
- Pulmonary edema
- Cardiogenic shock

0.8 % Probability of death from admission to 6 months

55 points GRACE Score

About the Creator

Dr. Joel Gore

Dr. Keith A. A. Fox

Also from MDCalc...

Related Calcs

- Framingham Risk Score (Hard Coronary Heart Disease)
- CRUSADE Score
- TIMI Risk Index

You might be interested in...

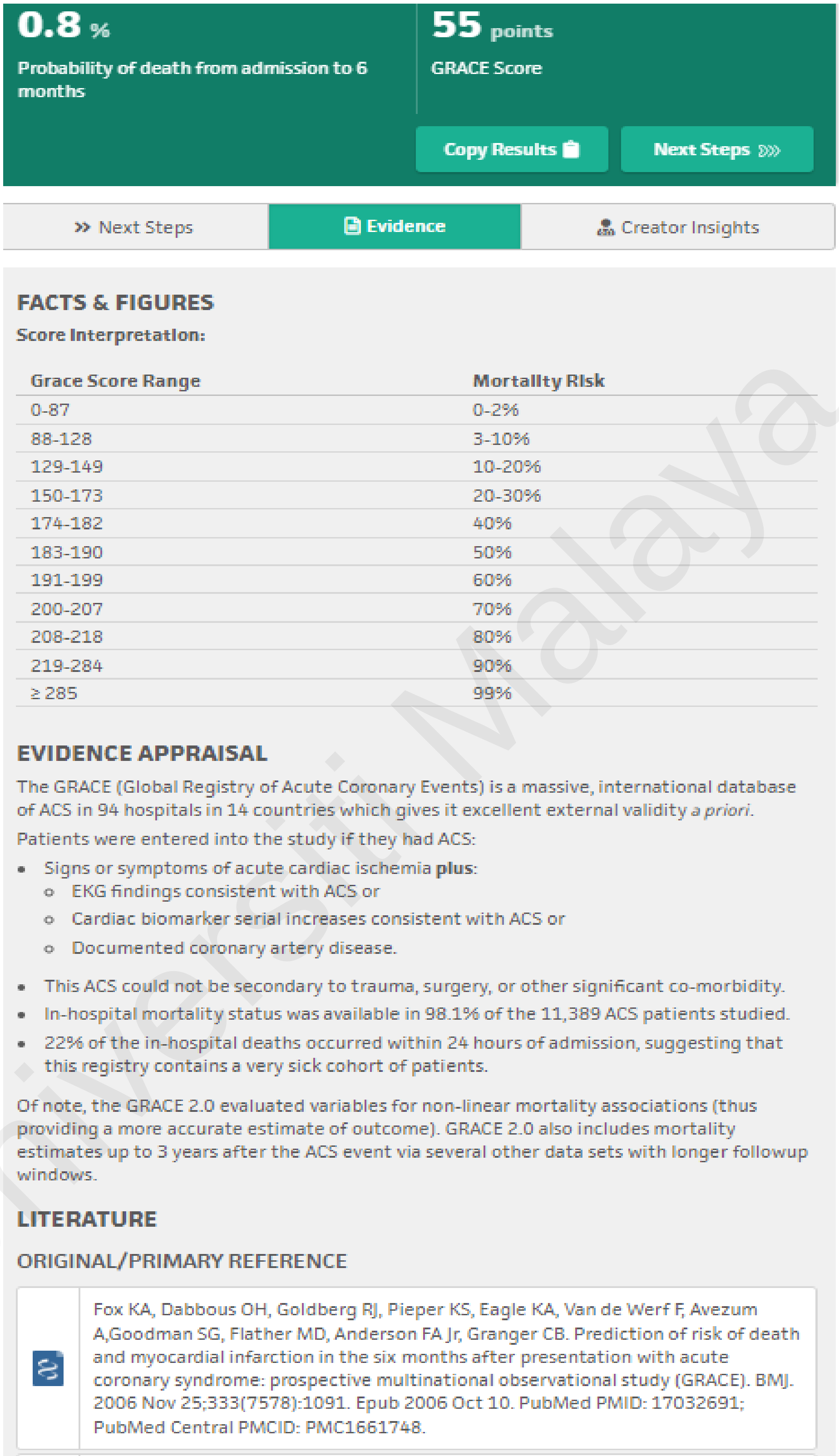
Partner Content

Calculated Decisions: GRACE ACS Score
Emergency Medicine Practice

Content Contributors

- Graham Walker, MD

Figure 2.21: Screenshot of MDCalc GRACE ACS risk and mortality calculator.



0.8 %

Probability of death from admission to 6 months

55 points

GRACE Score

FACTS & FIGURES

Score Interpretation:

Grace Score Range	Mortality Risk
0-87	0-2%
88-128	3-10%
129-149	10-20%
150-173	20-30%
174-182	40%
183-190	50%
191-199	60%
200-207	70%
208-218	80%
219-284	90%
≥ 285	99%

EVIDENCE APPRAISAL

The GRACE (Global Registry of Acute Coronary Events) is a massive, international database of ACS in 94 hospitals in 14 countries which gives it excellent external validity *a priori*.

Patients were entered into the study if they had ACS:

- Signs or symptoms of acute cardiac ischemia **plus**:
  - EKG findings consistent with ACS or
  - Cardiac biomarker serial increases consistent with ACS or
  - Documented coronary artery disease.
- This ACS could not be secondary to trauma, surgery, or other significant co-morbidity.
- In-hospital mortality status was available in 98.1% of the 11,389 ACS patients studied.
- 22% of the in-hospital deaths occurred within 24 hours of admission, suggesting that this registry contains a very sick cohort of patients.

Of note, the GRACE 2.0 evaluated variables for non-linear mortality associations (thus providing a more accurate estimate of outcome). GRACE 2.0 also includes mortality estimates up to 3 years after the ACS event via several other data sets with longer followup windows.

LITERATURE

ORIGINAL/PRIMARY REFERENCE

Fox KA, Dabbous OH, Goldberg RJ, Pieper KS, Eagle KA, Van de Werf F, Avezum A,Goodman SG, Flather MD, Anderson FA Jr, Granger CB. Prediction of risk of death and myocardial infarction in the six months after presentation with acute coronary syndrome: prospective multinational observational study (GRACE). BMJ. 2006 Nov 25;333(7578):1091. Epub 2006 Oct 10. PubMed PMID: 17032691; PubMed Central PMCID: PMC1661748.

Figure 2.22: Result generated of the GRACE Score along with the probability of death from admission to 6 Months.

2.12.1.3 OMNI Health Calculator for TIMI (STEMI) and TIMI (NSTEMI/UA)

OMNI Calculator is a Polish founded startup and developed custom built calculators for everyday use to solve mathematical problem. Part of it, is the OMNI Health Calculator, which is a web-based tool that consist of various health calculators, like body measurements calculator, dietary calculators, urology and nephrology calculators, cardiovascular calculators, etc. comes with explanation of medical terms and assist in calculating the result.

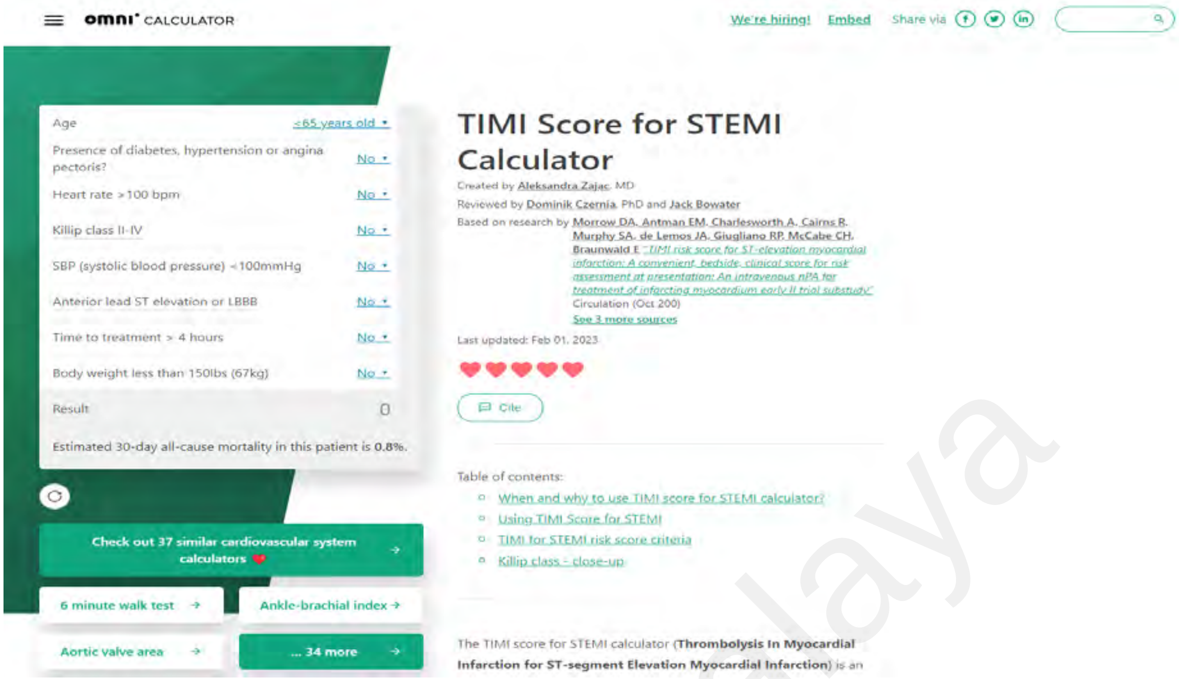
There are two versions of the TIMI score available on the OMNI Health Calculator under the cardiovascular system calculators, since the score calculated for STEMI and UA/NSTEMI patients are different:

(a) TIMI Score Calculator for STEMI Calculator

The calculator predicts 30-day all-cause mortality for STEMI patients, created by Aleksandra Zajac, it is solely developed based on Morrow, et al. (2000) study. The calculator comes with the drop-down list box, then the user is required to answer "Yes" or "No", moving on to the next criteria. The result will display automatically after the user fill in the final field (Zając, 2023).

The calculator has an interface and comes with explanation of the calculator, including "When and why to use the TIMI score for STEMI calculator", "Using TIMI Score for STEMI", "TIMI Score for STEMI risk score criteria", "Killip class – close-up". Figure 2.23 and figure 2.24 are the screenshots of the website and the URL of the website is given below:

URL: https://www.omnicalculator.com/health/timi-stemi



Age	≤65 years old
Presence of diabetes, hypertension or angina pectoris?	No
Heart rate >100 bpm	No
Killip class II-IV	No
SBP (systolic blood pressure) <100mmHg	No
Anterior lead ST elevation or LBBB	No
Time to treatment > 4 hours	No
Body weight less than 150lbs (67kg)	No
Result	0

Estimated 30-day all-cause mortality in this patient is **0.8%**.

TIMI Score for STEMI Calculator

Created by Aleksandra Zajac, MD

Reviewed by Dominik Czernia, PhD and Jack Bowater

Based on research by Morrow DA, Antman EM, Charlesworth A, Cairns R, Murphy SA, de Lemos JA, Giugliano RP, McCabe CH, Braunwald E *"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 (Oct 200)

See 3 more sources

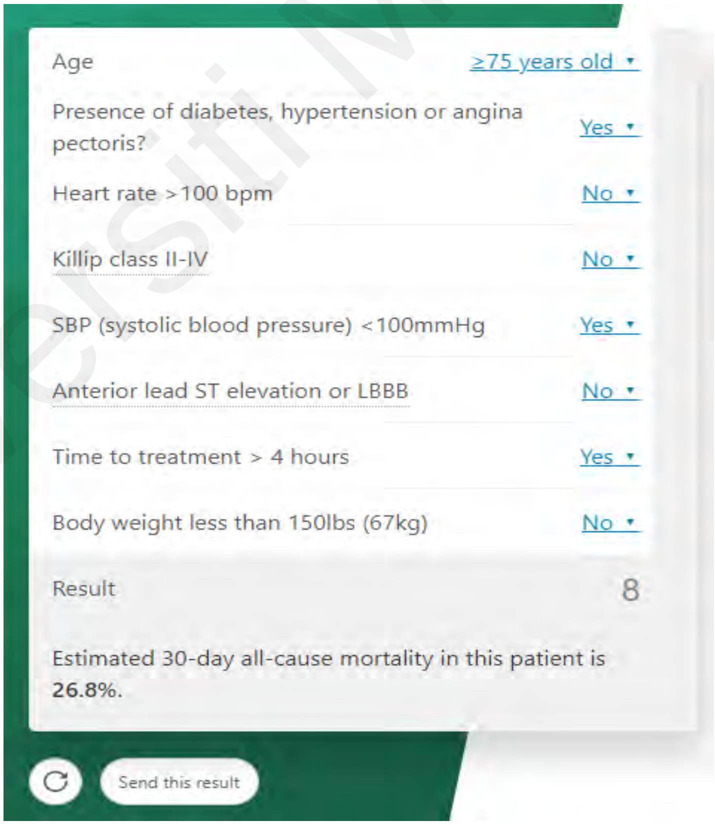
Last updated: Feb 01, 2023

Table of contents:

- When and why to use TIMI score for STEMI calculator?
- Using TIMI Score for STEMI
- TIMI for STEMI risk score criteria
- Killip class - close-up

The TIMI score for STEMI calculator (**Thrombolysis In Myocardial Infarction for ST-segment Elevation Myocardial Infarction**) is an

Figure 2.23: OMNI TIMI score for STEMI Calculator.



Age	≥75 years old
Presence of diabetes, hypertension or angina pectoris?	Yes
Heart rate >100 bpm	No
Killip class II-IV	No
SBP (systolic blood pressure) <100mmHg	Yes
Anterior lead ST elevation or LBBB	No
Time to treatment > 4 hours	Yes
Body weight less than 150lbs (67kg)	No
Result	8

Estimated 30-day all-cause mortality in this patient is **26.8%**.

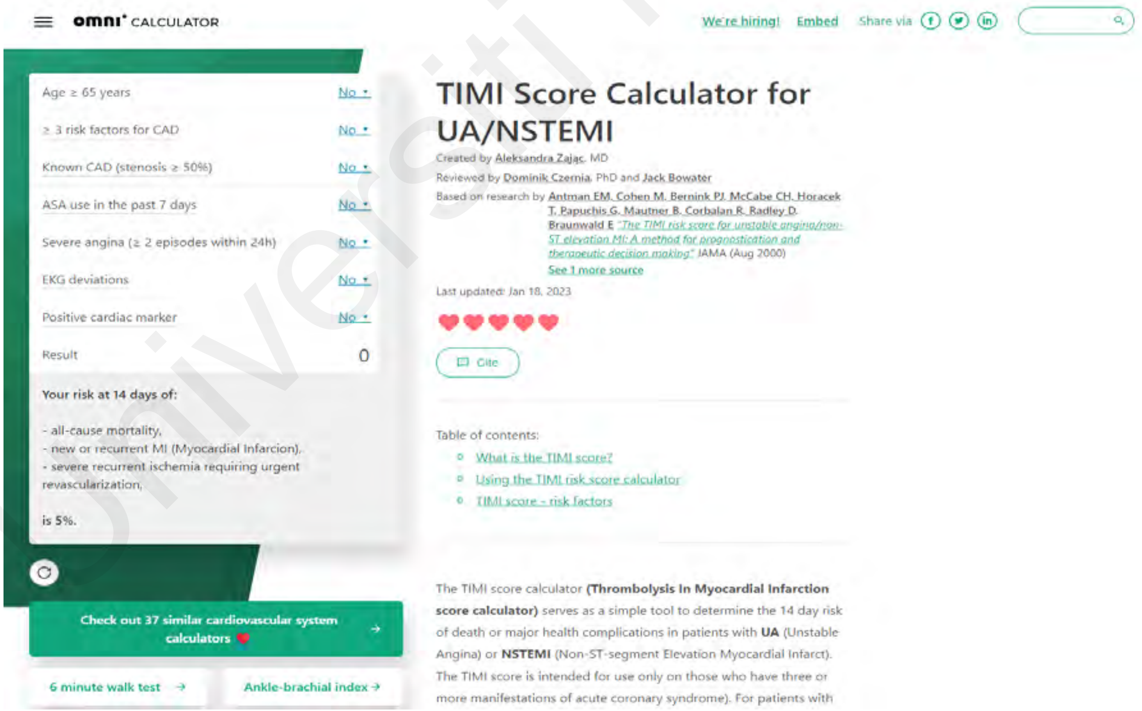
Figure 2.24: The result shown the estimated 30-Day-all-cause mortality for STEMI patient based on TIMI risk score criteria.

(b) TIMI Score Calculator for UA/NSTEMI Calculator

This score calculator is a simple tool meant for patients with UA or NSTEMI to determine the 14 days risk of death or major health complications (Zając, TIMI Score Calculator for UA/NSTEMI, 2023b). It was created by Aleksandra Zajac, based on the study by (Antman, et al., 2000). The research has shown that TIMI score correlates with the risk of an adverse outcome and is a valuable and simple prognostic tool.

The interface is similar to TIMI score calculator for STEMI patients as mentioned at part (a), the differences are the risk score factors, and the result estimates the patient risk at 14 days. Figure 2.25 and 2.26 shows the print screen of the TIMI score calculator for UA/NSTEMI Calculator.

URL: https://www.omnicalculator.com/health/timi-ua-nstemi



Age ≥ 65 years	No
≥ 3 risk factors for CAD	No
Known CAD (stenosis ≥ 50%)	No
ASA use in the past 7 days	No
Severe angina (≥ 2 episodes within 24h)	No
EKG deviations	No
Positive cardiac marker	No
Result	0

Your risk at 14 days of:

- all-cause mortality,
- new or recurrent MI (Myocardial Infarcion),
- severe recurrent ischemia requiring urgent revascularization,

is 5%.

TIMI Score Calculator for UA/NSTEMI

Created by Aleksandra Zając, MD
Reviewed by Dominik Czernia, PhD and Jack Bowater
Based on research by Antman EM, Cohen M, Bernink PJ, McCabe CH, Horacek T, Papuchis G, Mautner B, Corbalan R, Radley D, Braunwald E *"The TIMI risk score for unstable angina/non-ST elevation MI: A method for prognostication and therapeutic decision making"* JAMA (Aug 2000)
See 1 more source

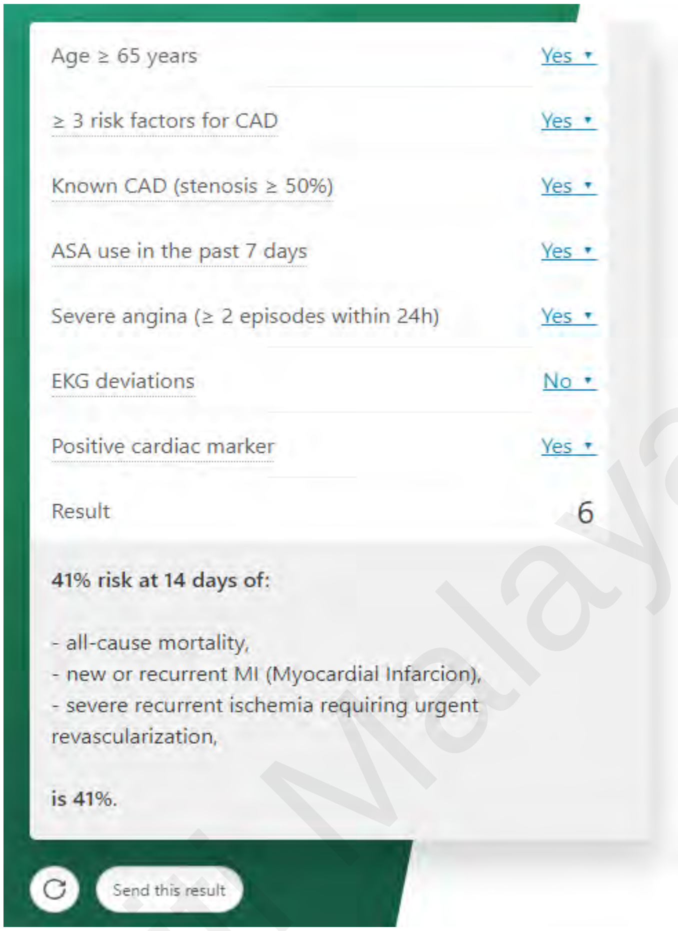
Last updated: Jan 18, 2023

Table of contents:

- What is the TIMI score?
- Using the TIMI risk score calculator
- TIMI score - risk factors

The TIMI score calculator **(Thrombolysis In Myocardial Infarction score calculator)** serves as a simple tool to determine the 14 day risk of death or major health complications in patients with **UA** (Unstable Angina) or **NSTEMI** (Non-ST-segment Elevation Myocardial Infarct). The TIMI score is intended for use only on those who have three or more manifestations of acute coronary syndrome). For patients with

Figure 2.25: OMNI TIMI score for UA/NSTEMI Calculator.



Age ≥ 65 years Yes

≥ 3 risk factors for CAD Yes

Known CAD (stenosis ≥ 50%) Yes

ASA use in the past 7 days Yes

Severe angina (≥ 2 episodes within 24h) Yes

EKG deviations No

Positive cardiac marker Yes

Result 6

41% risk at 14 days of:

- all-cause mortality,
- new or recurrent MI (Myocardial Infarcion),
- severe recurrent ischemia requiring urgent revascularization,

is 41%.

Figure 2.26: The result shown the estimated 30-Day-all-cause mortality for UA/NSTEMI patient based on TIMI risk score criteria.

2.12.2 Existing Air Quality Monitoring Web Application

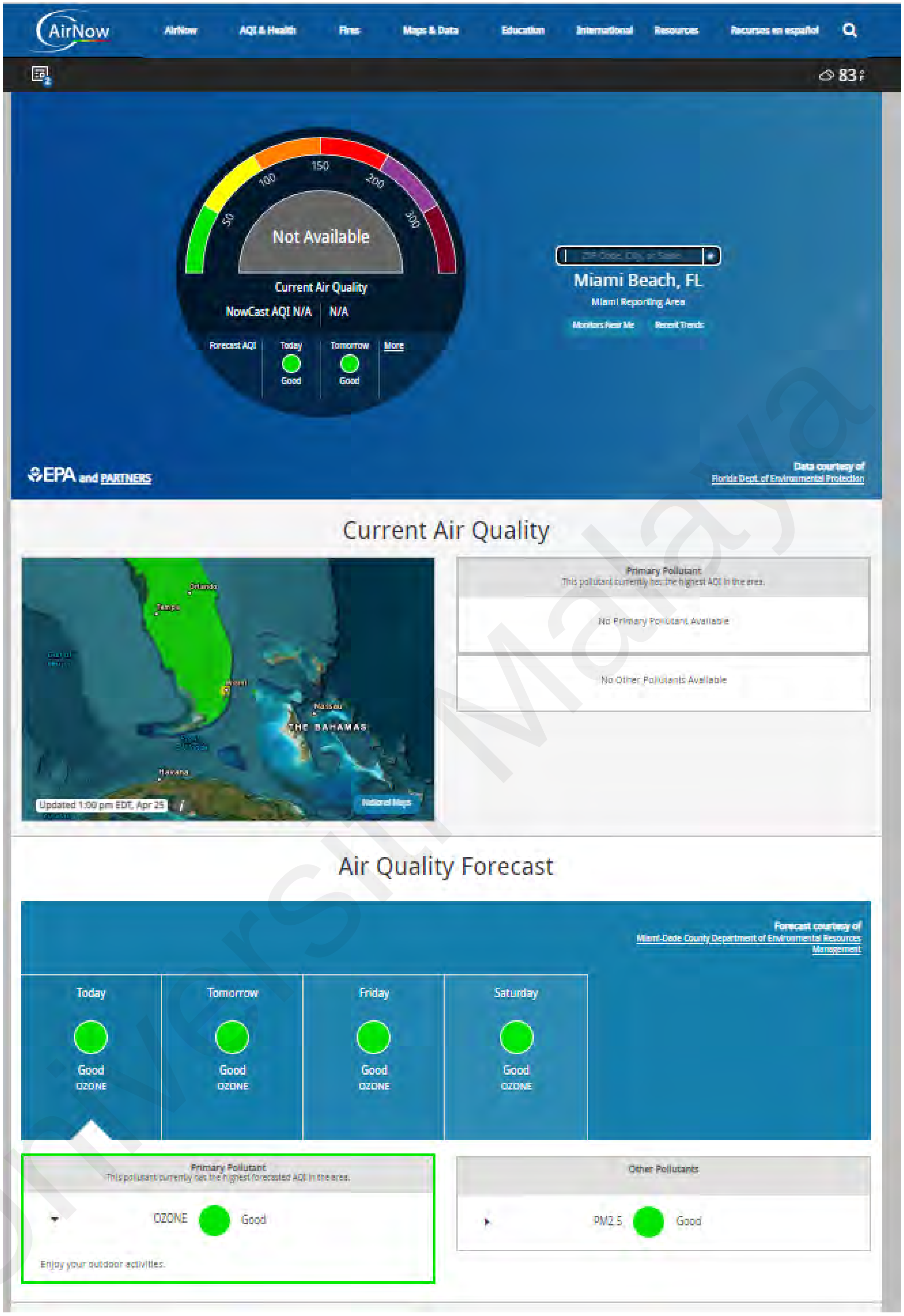
According to the findings of numerous studies, air pollutants interfere with various human physiological systems, which has a negative impact not only on health but also on climate change. human capital and the economy on a global scale (Giri, et al., 2023; Fisher, et al., 2021; Liao, et al., 2021; Manisalidis, et al., 2020; De Marco, et al., 2019). Air pollution and its detrimental impact on human health have become a growing concern for public health officials and policymakers around the world. To address this issue, several air monitoring systems that track and report on pollution levels in various geographic regions have been developed. There are several existing online air monitoring systems, including AirNow, World Air Pollution Index, and BreezoMeter.

2.12.2.1 AirNow

AirNow is developed to provide the public with real-time data to allow them to make lifestyle decisions to reduce or avoid exposure to poor air quality (Dickerson, 2012). AirNow is a centralized data management center that receive real-time zone and particle pollution data from 500 U.S. cities and an informational website for source and facts regarding air quality data and air quality forecast comes along with suggestions. The site consists of an interactive map that allows users to interact to obtain a better overall perspective and view data for an individual air quality monitor. Besides, the website emphasizes local air quality and provides air quality information (White, et al., 2004).

AirNow is a US government-run web system that monitors air quality across the United States using the official U.S. Air Quality Index (AQI) which calculated using data from a 24-hour period applying the NowCast algorithm where it is designed to be responsive to rapidly changing air quality conditions. It primarily utilizes data from government monitoring stations to report on pollutant levels, including PM2.5, ozone, NO_2, and SO_2 focusing on the United States and Canada, other countries and region do not have information regarding the forecasted air quality. Figure 2.27 is the homepage of AirNow web application. To access the website, the website address is given below.

URL: https://www.airnow.gov/



Not Available

Current Air Quality

NowCast AQI N/A | N/A

Forecast AQI | Today ● Good | Tomorrow ● Good

Miami Beach, FL

Miami Reporting Area

Data courtesy of
Florida Dept. of Environmental Protection

Current Air Quality

Updated 1:00 pm EDT, Apr 25

Primary Pollutant
This pollutant currently has the highest AQI in the area.

No Primary Pollutant Available

No Other Pollutants Available

Air Quality Forecast

Forecast courtesy of
Miami-Dade County Department of Environmental Resources Management

Today	Tomorrow	Friday	Saturday
●	●	●	●
Good	Good	Good	Good
OZONE	OZONE	OZONE	OZONE

Primary Pollutant
This pollutant currently has the highest forecasted AQI in the area.

OZONE ● Good

Enjoy your outdoor activities.

Other Pollutants

PM2.5 ● Good

Figure 2.27: AirNow homepage that shows air quality in Miami beach along with air quality forecast of the selected location.

2.12.2.2 World Air Quality Index

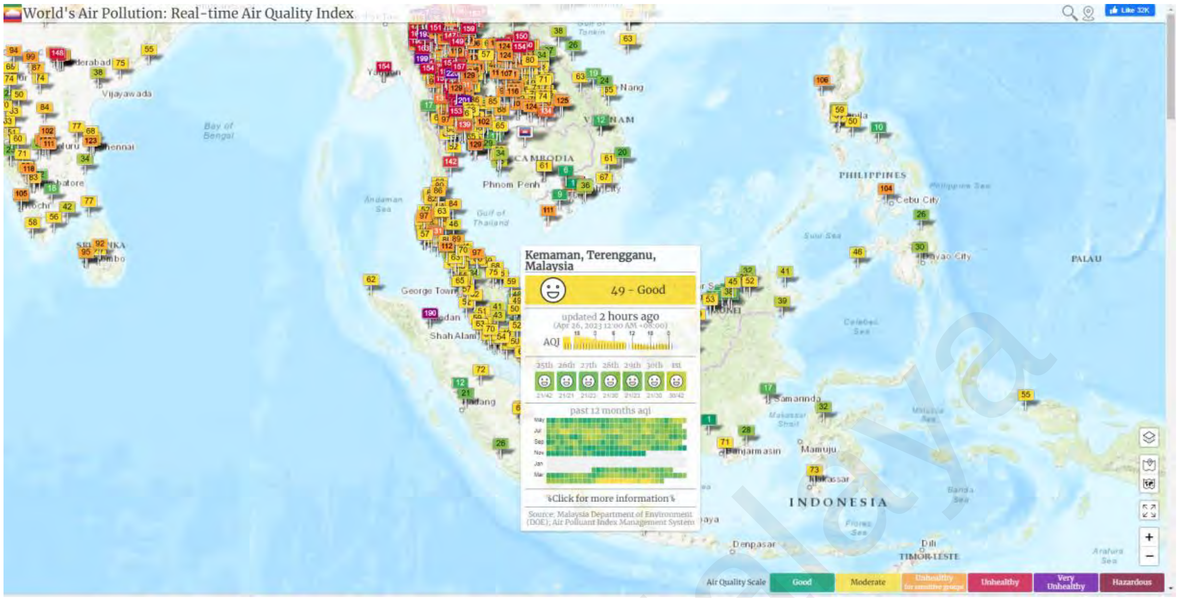
The World Air Quality Index is a web system that provides transparent air quality information for more than 130 countries, covering more than 30,000 stations in 2000 major cities, the data is collected and process via these websites: aqicn.org and waqi.info.

The World Air Quality Index project is a non-profit project started in 2007. Its mission is to promote air pollution awareness for citizens and provide unified and world-wide air quality information. This project is founded by several contributors in the domain environmental sciences, system engineering, data science, as well as visual design located in Beijing, China (WAQ, 2020).

The Air Quality Index is based on measurement of particulate matter (PM2.5 and PM10), Ozone (O_3), Nitrogen Dioxide (NO_2), Sulfur Dioxide (SO_2) and Carbon Monoxide (CO) emissions. Most of the stations on the map are monitoring both PM2.5 and PM10 data, but there are few exceptions where only PM10 is available. The calculated AQI is then categorized according to the air quality scale as shown in Figure 2.28. The colors of the flags are also shown on the map, that allows user to have an overall picture of which location has poor air quality. To get more information about a specific city, user is required to move over any of the flags in the above map, then click to get the full air pollution historical data shown in Figure 2.29 (WAQ, 2020).

All the Air Quality data seen on World Air Quality Index are the official data from each country respective Environmental Protection Agency (EPA). Data from each EPAs is measured using professional monitoring equipment, and only stations with particulate matter (PM10/PM2.5) readings are published. For Malaysia, though the system has the full coverage of the country, only pre-calculated AQI is provided. No detailed Individual pollutant AQI (IAQI) is available (WAQ, 2020).

URL: https://waqi.info/



World's Air Pollution: Real-time Air Quality Index

Kemaman, Terengganu, Malaysia

49 - Good

updated 2 hours ago
(Apr 26, 2023 12:00 AM +08:00)

AQI

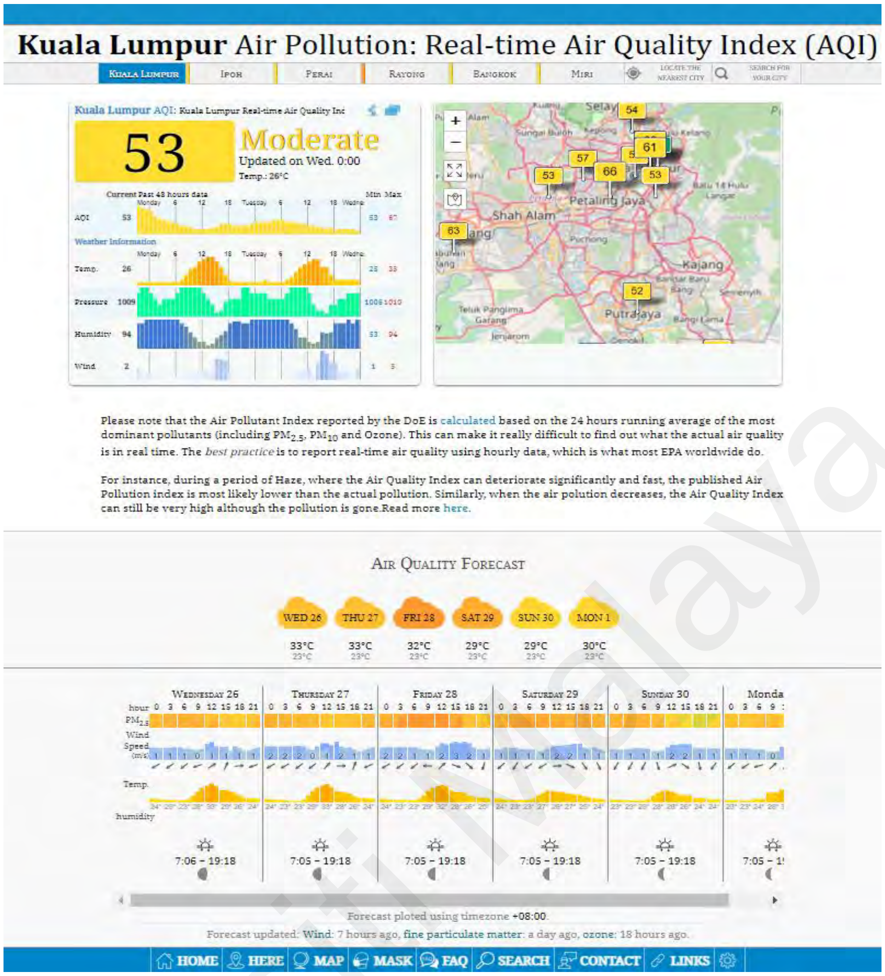
25th	26th	27th	28th	29th	30th	1st
21/42	21/21	21/22	21/30	21/22	21/30	30/42

past 12 months aqi

Click for more information

Source: Malaysia Department of Environment (DOE), Air Pollutant Index Management System

Figure 2.28: World Air Quality Index (WAQI) homepage.



Kuala Lumpur Air Pollution: Real-time Air Quality Index (AQI)

Please note that the Air Pollutant Index reported by the DoE is calculated based on the 24 hours running average of the most dominant pollutants (including $PM_{2.5}$, PM_{10} and Ozone). This can make it really difficult to find out what the actual air quality is in real time. The *best practice* is to report real-time air quality using hourly data, which is what most EPA worldwide do.

For instance, during a period of Haze, where the Air Quality Index can deteriorate significantly and fast, the published Air Pollution index is most likely lower than the actual pollution. Similarly, when the air polution decreases, the Air Quality Index can still be very high although the pollution is gone.Read more here.

Air Quality Forecast

Forecast ploted using timezone +08:00.

Forecast updated: Wind: 7 hours ago, fine particulate matter: a day ago, ozone: 18 hours ago.

Figure 2.29: Air Quality information once the user clicks on interested location on the interactive map.

2.12.2.3 Breezometer

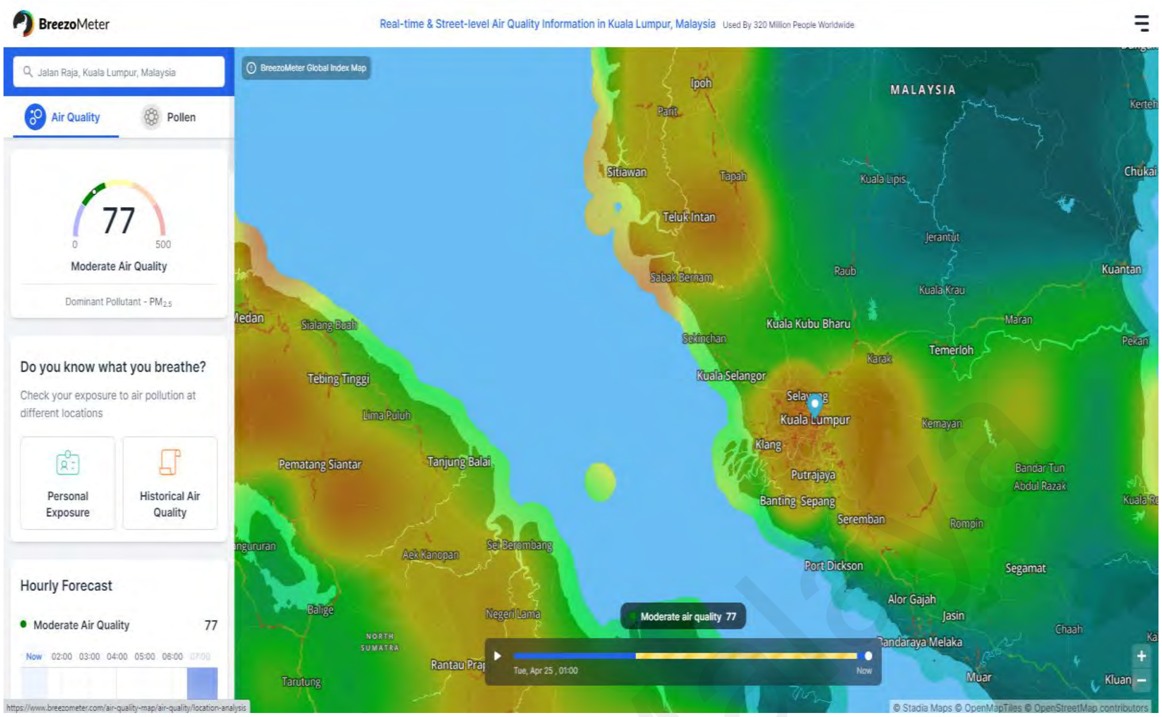
BreezoMeter is a web application that offers location-based, real-time environmental data on weather, wildfires, pollen, and air pollution covering more than 100 countries founded in 2014. BreezoMeter utilizes Google Cloud to support its comprehensive environmental intelligence platform, thereby making the invisible visible and mitigating the global effects of air pollution (Fisher & Korber, 2014).

BreezoMeter is a proprietary algorithm-based web system that tracks air quality globally, using data on pollutants including PM2.5, PM10, ozone, and NO_2. It offers health recommendations, pollen data, and a mobile app. However, it does not rely on government monitoring stations or provide historical data, as the accuracy of BreezoMeter is measure by developing a cross-validation model based on the "Leave one out" method, remove the data from one government sensor out of the script every hour, calculate, and compare it to what the sensor is reading. This method works continuously behind the scenes of our information, ensuring accuracy and early detection of irregularities (Breezometer, 2023).

On the BreezoMeter Real-time & Street-level Air Quality Information in Kuala Lumpur webpage, it shows an interactive map that allows users to view real-time air quality data at street-level resolution, by placing the marker on the map, on the sidebar, it will display information regarding pollutant levels, hourly forecast, health advice, and air pollution sources. The data is presented using a color-coded system, which indicates the level of air pollution in each area. The color scheme ranges from green (good air quality) to red (hazardous air quality).

The homepage of Breezometer comes with the interactive map as shown in Figure 2.30. By scrolling down the sidebar, a detailed measurement on the air pollutant (CO, NOx, SO_2, O_3 and PMs) are displayed as shown in Figure 2.31 below.

URL: https://www.breezometer.com/air-quality-map/air-quality/



Real-time & Street-level Air Quality Information in Kuala Lumpur, Malaysia Used By 320 Million People Worldwide

Jalan Raja, Kuala Lumpur, Malaysia

77 Moderate Air Quality

Dominant Pollutant - $PM_{2.5}$

Do you know what you breathe?

Check your exposure to air pollution at different locations

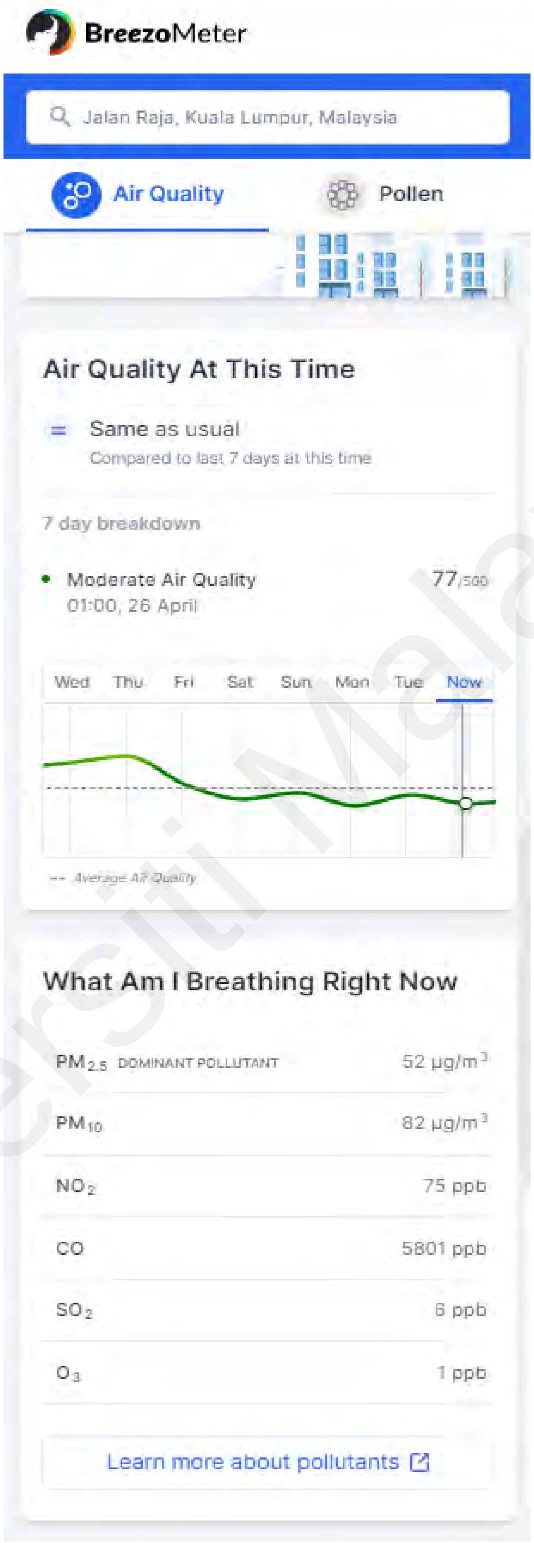
Hourly Forecast

Moderate Air Quality 77

Moderate air quality 77

Tue, Apr 25 , 01:00

Figure 2.30: Breezometer homepage along with air pollution interactive map.



Air Quality At This Time

= Same as usual

Compared to last 7 days at this time

7 day breakdown

● Moderate Air Quality 77/500

01:00, 26 April

What Am I Breathing Right Now

$PM_{2.5}$ DOMINANT POLLUTANT	52 µg/m^3
PM_{10}	82 µg/m^3
NO_2	75 ppb
CO	5801 ppb
SO_2	6 ppb
O_3	1 ppb

Figure 2.31: Air pollutants readings on Breezometer.

2.12.3 Comparison of Existing Web Application

Table 2.17 below summarized and compared the web application discussed previously:

Table 2.17: Overview of the existing mortality prediction calculator and air quality monitoring web system.

Web Application / Systems	Purpose and Functionality	User Interaction	Geographic Coverage	Methods Used / Data source	Availability	Data Visualization	Drawbacks
Cleveland Clinic Risk Calculator	Predicts 30-day and 1-year risks of mortality, myocardial infarction, or revascularization for ACS patients recently discharged	Minimal	Worldwide	Logistic Regression	Free and available to public	Risk of mortality, myocardial infraction, and revascularization for 30 days and 1 year.	Does not include visualization element. Environmental factors are not considered.
MDCalc for GRACE ACS Risk and Mortality Calculator	Estimates admission-6 months mortality for patients with acute coronary syndrome.	Minimal	Worldwide	GRACE risk score	Free and available to public	Simple risk score and mortality prediction based on GRACE risk score.	Does not include visualization element. Environmental factors are not considered.

Table 2.17, continued.

Web Application / Systems	Purpose and Functionality	User Interaction	Geographic Coverage	Methods Used / Data source	Availability	Data Visualization	Drawbacks
OMNI Health Calculator for TIMI (STEMI) and TIMI (NSTEMI/UA)	To predict mortality and cardiovascular risk in ACS patients	Moderate	Worldwide	TIMI risk score for STEMI and UA/NSTEMI	Free and available to public	Simple risk score and mortality prediction on TIMI risk score.	Does not include visualization element. Environmental factors are not considered.
AirNOW	Provide real-time air quality information	Moderate	United States	Environmental Protection Agency (EPA)	Free and available to public	Current air quality reading with non-interactive map, and air quality forecast of primary pollutants.	Limited to United states region. Does not include prediction element.
World Air Quality Index (WAQI)	Provide real-time air quality information	Moderate	Worldwide	Various government and private organizations	Free and available to public	Interactive map with real-time air quality index and colored markers that indicate the air quality.	Does not include prediction element.

Table 2.17, continued.

Web Application / Systems	Purpose and Functionality	User Interaction	Geographic Coverage	Methods Used / Data source	Availability	Data Visualization	Drawbacks
Breezometer	Provide real-time and street-level air quality information	Easy	Worldwide	Multiple data sources, including satellites	Free and available to public	Interactive map that shows real-time and street-level air quality information with color scheme ranges from green to red.	Does not include prediction element.

2.13 Summary of Literature Review

In conclusion, this review of the literature has demonstrated the significant impact that air pollution can have on the incidence of ACS. Several key air pollution features that are most strongly associated with the onset of ACS were identified through an extensive review of previous research. This chapter discussed the potential of ML models for predicting ACS hospitalization and mortality rates in the presence of air pollution, as well as predicting the risk of mortality in ACS patients. Furthermore, the conventional risk scoring method was presented and compared to existing ML studies in predicting the mortality rate of ACS patients.

Several significant gaps in the literature were identified, including the need for additional research on the effects of air pollution exposure on ACS incidence through ML approach, as well as the potential impact of air pollution. More research on the development of web applications and visualization tools for presenting air pollution and health data to stakeholders is required. In addition, using ML to predict the occurrence of ACS in the presence of air pollution is vital for the Southeast Asia population.

Overall, this review lays the foundation for future research into the relationship between air pollution and ACS, highlighting the potential of ML models and visualization tools to improve our understanding of this public health issue. The findings of this study are expected to be useful to healthcare professionals, policymakers, and other stakeholders working to improve air quality and public health in Malaysia and Southeast Asia.

CHAPTER 3: MATERIALS AND METHODS

3.1 Introduction

This chapter summarizes the materials and research methodologies used in this study. The primary goal of this study is to determine the effect of air pollution on hospitalization and mortality rates in acute coronary syndrome (ACS) patients. Furthermore, the objective of this study is to develop a mortality risk prediction calculator in the presence of air pollution using machine learning (ML) and stacked ensemble learning (EL) approach. The best performing ML models are integrated into an interactive web system with visualization features for user understanding and interaction.

This chapter is divided into several sections that outline the study design, study data and data preprocessing procedures, ML algorithms applied, and web system development process was designed to achieve the research objectives. The flowchart of the research process is depicted in Figure 3.1 below.

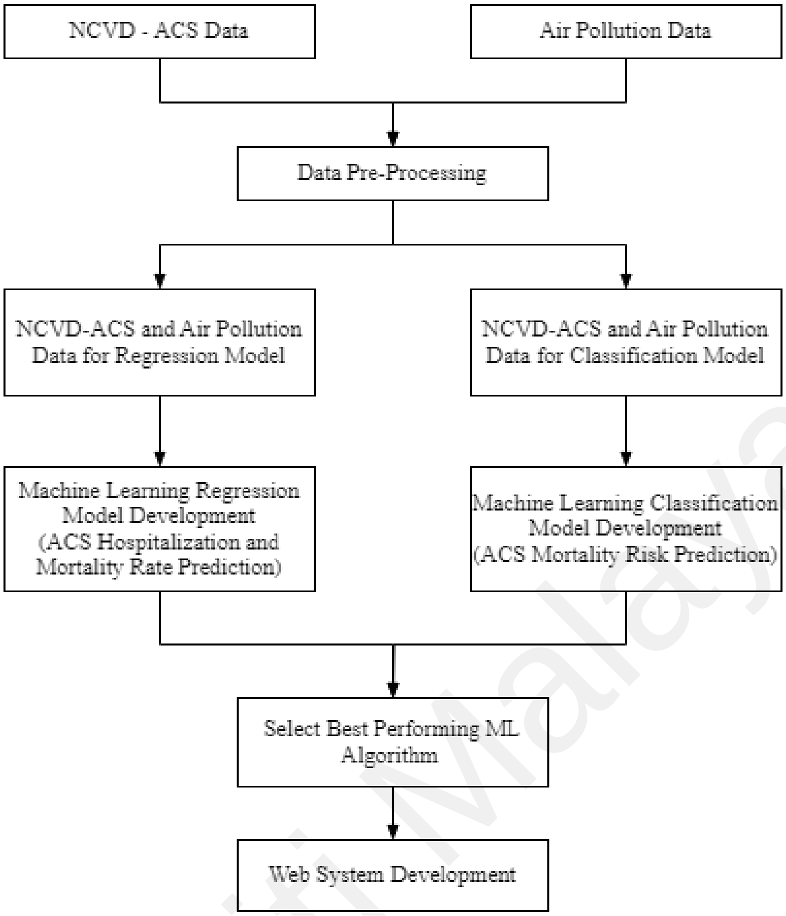

NCVD - ACS Data
Air Pollution Data
Data Pre-Processing
NCVD-ACS and Air Pollution Data for Regression Model
NCVD-ACS and Air Pollution Data for Classification Model
Machine Learning Regression Model Development (ACS Hospitalization and Mortality Rate Prediction)
Machine Learning Classification Model Development (ACS Mortality Risk Prediction)
Select Best Performing ML Algorithm
Web System Development

Figure 3.1: General flowchart of the study.

3.2 Study Data

The data for this study was collected from two primary sources: The National Cardiovascular Disease Database (NCVD) for ACS data and the Department of Environment (DOE), Malaysia for air quality data. Both datasets were received as structured data.

1. Acute Coronary Syndrome (ACS) data

The National Cardiovascular Disease Database (NCVD) is a service supported by the Ministry of Health (MOH) to collect information about cardiovascular disease in Malaysia, enabling the determination of the incidence of cardiovascular disease (CVD) in the country. The national cardiovascular disease database (NCVD-ACS) registry data will be used from 2006 – 2017. NCVD-ACS is a collaborative multicenter registry involving 25 hospitals across Malaysia (Appendix A). The registry collects data on a standardized set of clinical, demographic, and procedural variables, along with outcomes, for consecutive patients treated at participating institutions.

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. The UiTM ethics committee (Reference number: 600-TNCPI (5/1/6)) and the National Heart Association of Malaysia (NHAM) both granted their approval for data collection. Deaths were confirmed on a yearly basis through record connections with the Malaysian National Registration Department of Deaths.

Data were collected using a standardized 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 (Ahmad, et al., 2011). A copy of the case report is attached in Appendix B.

The hospital cardiologist decided the ACS diagnosis based on clinical symptoms, electrocardiogram as well as cardiac biomarkers. The initial 54 variables were selected by the cardiologist. In this study, 14 features were selected from the NCVD registry. The features are patient’s age, heart rate, ECG abnormalities past 2 weeks, cardiac catheterization, coronary artery bypass graft

(CABG), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), fasting blood glucose (FBG), Killip class, chronic angina, intake of statin, oral hypoglycemic agent, anti-arrhythmic and lipid lowering agent medications.

The classification model in this study was based on selected features from previous study (Kassim et al., 2022). Using a similar registry dataset NCVD, Kasim et al. (2022) found that only 14 SVM features with deep learning classifier features are highly associated with ACS mortality (refer to Appendix L). Thus, ACS patient mortality using these 14 ACS variables and air pollution parameters is studied.

2. Air quality data

The air quality data for this study was obtained from the Department of Environment (DOE), Malaysia, covering the period between 1st January 2006 and 13th April 2017. The dataset includes daily air quality measurements for key air pollutants such as Nitrogen Oxides (NOx), Sulphur Dioxide (SO_2), Ozone (O_3), and Particulate Matter 10 (PM10). It is important to highlight that PM2.5 statistics were not available during this timeframe, due to unavailable technical resources by DOE, Malaysia.

A total of 61,816 instances of air quality data were collected during this period. This data was further processed, subsequently merged with NCVD-ACS data based on the geographical location of the hospitals, specifically within a 15 km radius (Khir, et al., 2018), covering an area of 706.85 km^2. To further support our findings, the locations of the monitoring stations and hospitals were plotted and analyzed using Google Earth. This step enhanced our understanding of the spatial relationship between these entities and aided in the accurate merging of the air quality and NCVD-ACS datasets. For a comprehensive view of the mapped locations, please refer to the images included in Appendix C.

DOE operates a network of air quality monitoring stations across Malaysia to provide representative measurements of ambient air quality. The dataset used in this study comprises the 24-hour mean concentrations for each of the pollutants, accounting for daily variations in pollutant levels.

The data has been divided into four levels of exposure to air pollution: lag 0 represents daily exposure, lag 03 represents exposure three days before the event, lag 07 represents average weekly exposure, and lag 30 represents monthly exposure. The time lag 00 is used in the classification model to predict mortality risk of ACS patients, as are the four time-lags are applied in the regression model to study the ACS hospitalization and mortality in the presence of air pollution.

3.3 Research Design Overview

To develop a visualization tool integrated into Google Map for easy access and understanding of the data, there are four main stages. Firstly, the data will be collected from various sources, such as air quality monitoring stations, hospital records, and demographic data of the population. Then, the ML algorithms will be used to analyze the data and determine the correlation between air pollution and the onset of ACS. The results of this study will provide valuable insights into the impact of air pollution on public health in Malaysia and can be used to inform policy decisions and interventions to mitigate the effects of air pollution on cardiovascular health.

Below is a brief explanation of each phase of this study.

1. Data Preparation and Requirements: In this phase, ACS in-hospital data provided by the NCVD registry, while the air quality data is obtained from the DOE, Malaysia. The data will be cleaned, processed, and analyzed to ensure its suitability for the study. This phase also involves identifying and obtaining any additional data needed for the research.

   (a) Data Preparation for Regression Model: The hospitalization rate and mortality rate are acquired and derived from NCVD-ACS cohort with the air pollution data arranged by 4 different time lags. This allowed us to explore both immediate and lagged relationships between air pollution exposure and ACS-related outcomes.

   (b) Data Preparation for classification: Two separate dataset is prepared for the classification models to enhance the specificity, as follows:

   a. In-Hospital Selected Features: 14 selected features from previous study (Kasim, et. al., 2022) that shows significance with ACS mortality risk is merged with lag 00 (daily) exposure which reflects the immediate effect of air pollution.

   b. Emergency Selected Features: The dataset is further reduced into 9 selected features by cardiologist for their relevance and utility in emergency settings, where invasive procedure and baseline investigation features were excluded due to the immediacy of emergency settings and combine with lag 0 (daily) air pollution exposure.

2. Machine Learning Algorithm Design and Development: This phase involves two types of ML algorithms design to analyze the data collected in the first phase. This

phase also includes testing and fine-tuning the algorithms to ensure their accuracy and effectiveness in analyzing the data.

   a. Regression Model Development: The algorithms will be used to determine the ACS hospitalization rate and mortality rate in the presence of air pollution are linear regression, SVM, RF, XGBoost and stacked EL (meta-learner: GLM).
   b. Classification Model Development: 6 ML algorithms are integrated to predict the mortality risk of ACS patients in the presence of air pollution, including: logistic regression, SVM, RF, XGBoost, Naïve Bayes and stacked EL (meta-learner: GLM).

3. Prototype Development on a Local Host: The third phase involves converting the best algorithm into a web system with geographical visualization. This phase includes implementing the ML model developed in the second phase into a web application, incorporating geographical visualization capabilities for a map-based data format. Additionally, new data will be obtained from the DOE and NCVD for model validation.
4. System Development, Testing, Conversion, and Evaluation: The fourth and final phase involves prototyping, testing of the web system and the ML model on a local host to ensure correct and effective functioning. Prototyping allows for refining the user interface and overall design of the web application, ensuring that it meets the needs of the end-users. Any bugs or issues will be identified and resolved during this phase. Furthermore, the web system will be converted and deployed on the Google Map platform for easy access and understanding of the data. The final stage of this phase system usability using SUS matrix.

The following flowchart in Figure 3.2 demonstrates the basic overview of the study project on the visualization and impact of air pollution on ACS onset in Malaysia. Each phase builds upon the previous one and is essential for the overall success of the study.

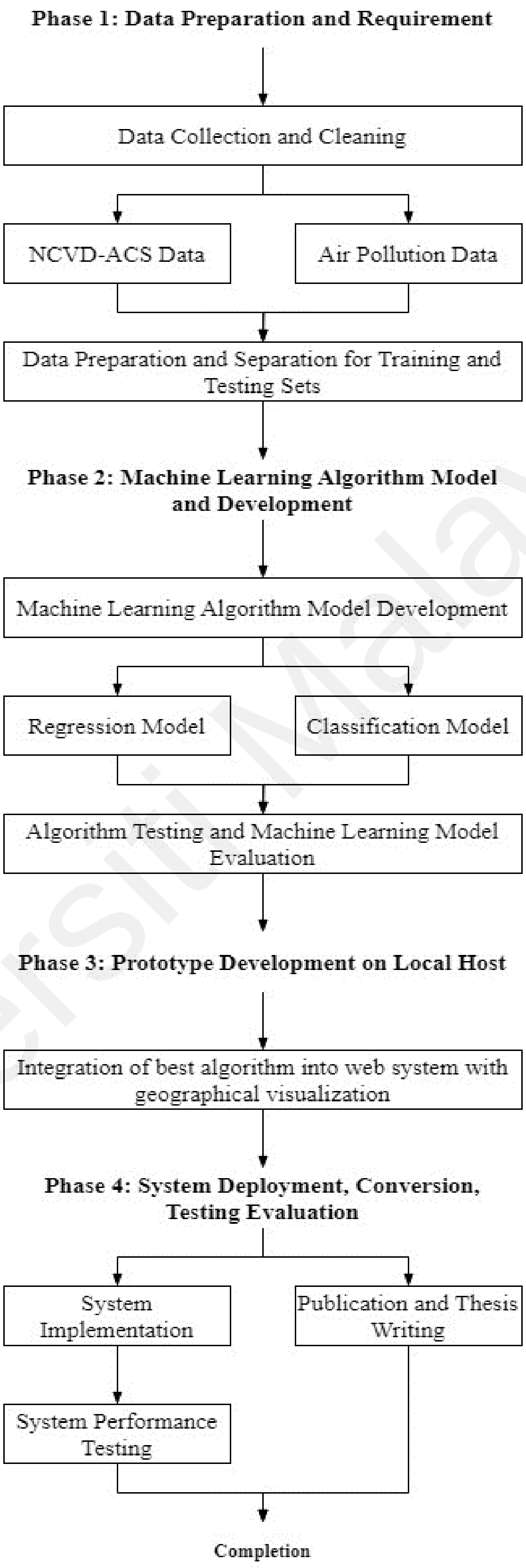

Phase 1: Data Preparation and Requirement
Data Collection and Cleaning
NCVD-ACS Data
Air Pollution Data
Data Preparation and Separation for Training and Testing Sets
Phase 2: Machine Learning Algorithm Model and Development
Machine Learning Algorithm Model Development
Regression Model
Classification Model
Algorithm Testing and Machine Learning Model Evaluation
Phase 3: Prototype Development on Local Host
Integration of best algorithm into web system with geographical visualization
Phase 4: System Deployment, Conversion, Testing Evaluation
System Implementation
Publication and Thesis Writing
System Performance Testing
Completion

Figure 3.2: Flowchart diagram illustrating the research methodologies employed in this study.

Figure 3.3 shows a detailed overview of Phase 2, which focuses on the development of ML algorithms that will be integrated into the web system.

ML development for this study consists of regression and classification models. The regression analysis aims to predict continuous outputs, specifically the ACS hospitalization rate and mortality rate, in relation to air pollution. On the other hand, the classification aspect focuses on determining the ACS mortality risk in the presence of air pollution. The performance of ML algorithms is evaluated, and the best performing model will then be incorporated into a web-based system.

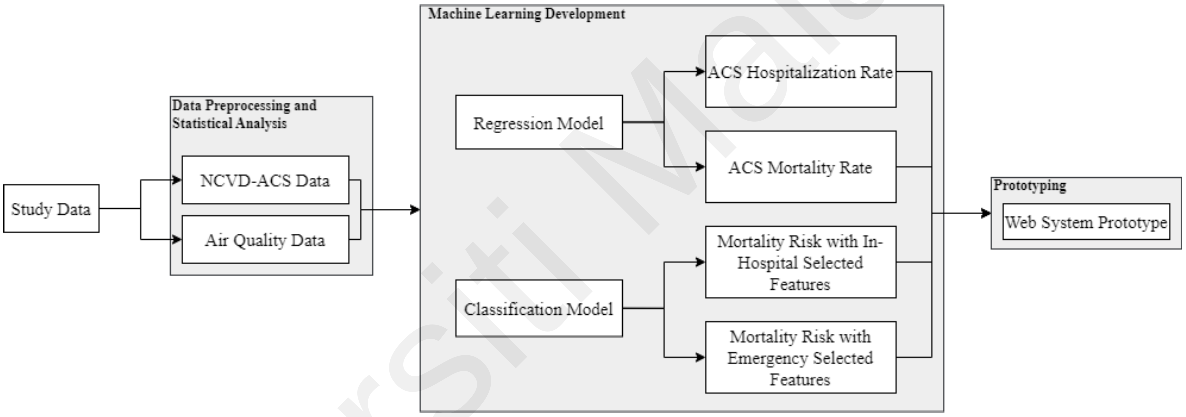

Study Data
Data Preprocessing and Statistical Analysis
NCVD-ACS Data
Air Quality Data
Machine Learning Development
Regression Model
ACS Hospitalization Rate
ACS Mortality Rate
Classification Model
Mortality Risk with In-Hospital Selected Features
Mortality Risk with Emergency Selected Features
Prototyping
Web System Prototype

Figure 3.3: Detailed summary of the study's second phase.

3.4 Data Pre-processing

Data cleaning, curation, and the removal of redundant features are carried out during the data preprocessing stage to ensure the data's quality and reliability. The preprocessing techniques used for regression and classification problems differ due to differences in input variables and expected outputs. The data is organized for future analysis using ML algorithms by adapting the preprocessing techniques to each problem type. The regression problem that predicts the number of ACS hospitalizations and mortality is examined in Section 3.4.1. The classification model used in this study, on the other hand, is described in Section 3.4.2.

3.4.1 Data Preprocessing for ML Regression

This study utilizes data from two different datasets: NCVD-ACS and DOE air quality data. Data from both sources were combined to examine the impact of air pollution on ACS patients. The regression model's goal is to predict the number of ACS hospitalizations and deaths among ACS patients in the presence of air pollution. In this study, air pollutant variables such as NOx, SO_2, O_3, and PM10 are examined in relation to hospitalization and mortality occurrences in ACS patients.

Previous research has demonstrated that both short-term and long-term exposure to air pollution can trigger the onset of ACS (Huynh, et al., 2018; Huynh, et al., 2021). Hence, four time-lags were used for the ML regression model development. There is no missing data for the air pollution data due to the daily collection of the air quality monitoring station. Lag00, representing daily exposure, consists of 57,694 data points. Lag03, which is arranged by exposure three days before the event, has two fewer data points. For lag07 and lag30, representing the average weekly and monthly data, there are 8,372 and 1,904 data points, respectively. Table 3.1 below summarizes the dataset for regression ML model that predicts hospitalization and mortality rate of ACS patients.

Table 3.1: Summary of ML Regression Analysis Dataset

	Lag00	**Lag03**	**Lag07**	**Lag30**
Dataset Characteristics	Multivariate	Multivariate	Multivariate	Multivariate
Number of Instances	57,694	57692	8,372	1,904
Attribute Characteristics	Integer, Real	Integer, Real	Integer, Real	Integer, Real
Number of Attributes	5	5	5	5
Associated Task	Regression	Regression	Regression	Regression
Missing Values	None	None	None	None

3.4.2 Data Preprocessing for Classification

In this section, the data preprocessing for the classification ML model is discussed in accordance with the research objective of identifying the impact of air pollution on ACS patients and developing a classification model that predicts ACS patient mortality risk in the presence of air pollution.

The raw dataset, obtained from the NCVD registry, comprises 54 variables. Two classification models are developed in this study: using 14 input features identified in a previous study by (Kasim, et al., 2022) and features used in emergency setting were further identified by cardiologist from the pre-selected 14 features.

This emergency dataset incorporates air quality variables but excludes baseline investigation features such as high-density lipoprotein (HDL-C), low-density lipoprotein (LDL-C), fasting blood glucose (FBG) levels, and invasive therapeutic procedures such as cardiac catheterization and coronary artery bypass graft (CABG). These characteristics required additional laboratory investigation, which is not available in emergency situations. For the development of the ML model, the selected cardiac features are combined with air pollution variables. The air pollution readings are based on time lag 00 (daily) readings because ACS onset is reported daily in the patient record, therefore corresponds to the patient record.

The merged datasets are examined for potential errors, missing values, or outliers, addressing them accordingly. The rows with incomplete data and outliers are removed, and only data with complete cases are retained. NCVD data with 54 variables total up to 54,000 records, since in this study, we have selected 14 variables this rendered in in-hospital dataset with 14,145 instances for model development. The emergency dataset with 22,466 instances, covering data from 2016 to 2017 for the entire Malaysian population. By removing these problematic rows, the quality and accuracy of the data used in the ML and EL model while minimizing the risk of introducing biases or inaccuracies (Psychogyios, et al., 2022). Table 3.2 shows the number of cases for selected variables and emergency variables, and Table 3.3 shows the percentage of missing values for the variables used in this study.

Table 3.2: Number of cases for selected variables and emergency variables.

	Selected Variables	**Emergency Variables**
Raw Data	50429	50429
Records with missing outcome	36284	27963
Data with complete cases	14145	22466

Table 3.3: The total and percentage of missing values in selected in-hospital variables and emergency variables.

Variables	**In-Hospital Variables**		**Emergency Variables**	
	Total missing value	**Percentage of missing value (%)**	**Total missing value**	**Percentage of missing value (%)**
Patient Age at Notification	0	0	0	0
Chronic Angina (>= 2 Weeks)	4998	9.91	4998	9.91
Heart Rate	1369	2.71	1369	2.71
Killip Class	12628	25.04	12628	25.04
High-density lipoprotein cholesterol (HDLC)	14006	27.77		
Low-density lipoprotein cholesterol (LDLC)	14055	27.87		
Fasting Blood Glucose	14034	27.83		
ST-segment elevation >= 1mm (0.1mV) in >= 2 contiguous limb leads*	0	0	0	0
Cardiac Catheterization	2421	4.80		
Coronary Artery Bypass Graft (CABG)	4085	8.10		
Statin	1803	3.58	1803	3.58
Other Lipid Lowering Agent	6113	12.12	6113	12.12

Table 3.3, continued.

Variables	In-Hospital Variables		Emergency Variables	
	Total missing value	Percentage of missing value (%)	Total missing value	Percentage of missing value (%)
Oral Hypoglycaemic Agent	5414	10.74	5414	10.74
Anti-arrhythmic Agent	6248	12.39	6248	12.39
Nitrogen Oxides	6702	13.29	6702	13.29
Sulphur Dioxide	6702	13.29	6702	13.29
Ozone	6702	13.29	6702	13.29
Particulate Matter 10	6702	13.29	6702	13.29

Table 3.4 displays the dataset summary with the in-hospital variables selected from the previous NCVD cohort study and the emergency variables combined with daily air pollutant exposure. Both datasets are arranged together with the air pollutants in time lag 00 (daily) value.

Table 3.4: Classification analysis dataset summary

	In-hospital Variables	Emergency Variables
Dataset Characteristics	Multivariate	Multivariate
Number of Instances	14145	22466
Attribute Characteristics	Integer, Real	Integer, Real
Number of Attributes	19	14
Associated Task	Classification	Classification
Missing Values	None	None

The summary statistics for both in-hospital patient and emergency Patient variables are presented in Table 3.5. The dataset has been cleaned and merged with air pollution data obtained from the NCVD Registry and the DOE, Malaysia. This integrated dataset serves as the study's foundation, providing important insights into the relationship between air pollution and ACS patient outcomes.

Table 3.5: Summary statistics for in-hospital selected and emergency patient with the time lag 00 exposure of air pollution.

Variables	Label	Data Domain	Mean	Std Dev.	Data Type
Demographic					
Patient Age at Notification*	ptageatnotification	19.89 – 101.94	58.2	12.14	Continuous
Status Before Event					
Chronic Angina (>= 2 Weeks) *	Canginapast2wk	0: No 1: Yes			Categorical
Clinical Presentation and Examination					
Heart Rate*	Heartrate	27 – 170 beats/min	83.18	20.42	Continuous
Killip Class*	Killipclass	1: Killip Class I 2: Killip Class II 3: Killip Class III 4: Killip Class IV			Categorical
Baseline Investigation					
High-density lipoprotein cholesterol (HDLC)	Hdlc	0.05 – 3.900 mmol/L	1.086	0.3174	Continuous
Low-density lipoprotein cholesterol (LDLC)	Ldlc	1.00 – 7.92 mmol/L	3.296	1.192	Continuous
Fasting Blood Glucose	Fbg	3.00 – 20.9 mmol/L	7.937	3.287	Continuous
Electrocardiography (ECG)					
ST-segment elevation >= 1mm (0.1mV) in >= 2 contiguous limb leads*	ecgabnormtypestelev1	0: No 1: Yes			Categorical
Invasive Therapeutic Procedure					
Cardiac Catheterization	Cardiaccath	0: No 1: Yes			Categorical
Coronary Artery Bypass Graft (CABG)	Cabg	0: No 1: Yes			Categorical

Table 3.5, continued.

Variables	Label	Data Domain	Mean	Std Dev.	Data Type
Pharmacological Therapy					
Statin*	Statin	0: No 1: Yes			Categorical
Other Lipid Lowering Agent*	Lipidla	0: No 1: Yes			Categorical
Oral Hypoglycaemic Agent*	Oralhypogly	0: No 1: Yes			Categorical
Anti-arrhythmic Agent*	Antiarr	0: No 1: Yes			Categorical
Air Pollutants					
Nitrogen Oxides*	nox	0 – 209.22 ppb	89.81	86.13	Continuous
Sulphur Dioxide*	so2	0 – 192.05 ppb	77.49	76.68	Continuous
Ozone*	o3	0 – 148.71 ppb	85.03	77.47	Continuous
Particulate Matter 10*	pm10	0 – 390 µg/m3	50.59	27.14	Continuous
In-Hospital Outcome					
Patient Outcome*	ptoutcome	0: non-survive 1: survive			Categorical

* Emergency Dataset Variables

3.5 Software Packages

The main language that was used throughout the study is R, where R packages offers a series of collections of R function which are stored under a directory called “library” in R (Harvard Chan Bioinformatics Core, 2023). Processes such as data pre-processing, data normalization, data balancing and the ML models’ development are performed using R. The ML models were developed with R package version 3.3.0.

Python was used to generate SHAP plots that provides further insight of the ML model. The python version used in this study is Python 3.10. Table 3.6 summarizes the libraries and packages that were used in this study.

Table 3.6: R and Python libraries used in this study.

Library/Packages	Functions
R	
caret	The acronym stands for Classification and Regression Training. ML tools for training regression and classification models, including pre-processing, training, tuning, and evaluating predictive models.
mlbench	Testing and comparing different ML algorithms.
pROC	Use for visualizing, smoothing, and comparing the ROC curves.
rstudioapi	Set working directory automatically.
plumber	Transform the developed ML model into web services. Integrates ML model with other applications.
dplyr	Simplify data manipulation and transformation data.
rose	The acronym stands for Random-Over Sampling Examples. Use to balance class distribution in binary distribution classification tasks.
ggplot2	Data visualization task, by creating graphs and charts.
caretEnsemble	Create ML model ensembles by combining the predictions of multiple models to improve the overall performances.
Python	
sklearn	The acronym stands for Scikit-learn. Data analysis and ML tasks including classification and regression.
numpy	The acronym stands for Numerical Python. Mathematical operations on arrays.
pandas	Data manipulation, such as data cleaning, data transformation and visualization.
matplotlib	Data visualization task, use for creating line plots, histograms of the import data.
shap	Interpret ML model using Shapley values, which gives an overall context of the ML model.

3.6 Data Partitioning

Following the data pre-processing stage, the cleaned data is organized and prepped for use in both ML regression and classification models. Data partitioning, or known as data splitting, is performed. 70% of the data was allocated for model training, and the remaining 30% was reserved for model testing as shown in Figure 3.4. The 7:3 ratio for data splitting is a common practice and widely accepted in ML practice, since it provides adequate data for model training, allowing the model to discover the data's underlying patterns and relationships (Hastie, et. al., 2009). From the raw data to the finalized training and testing data used in each of the model development, Figure 3.4 summarizes the data cleaning process for both in-hospital selected variables and emergency variables.

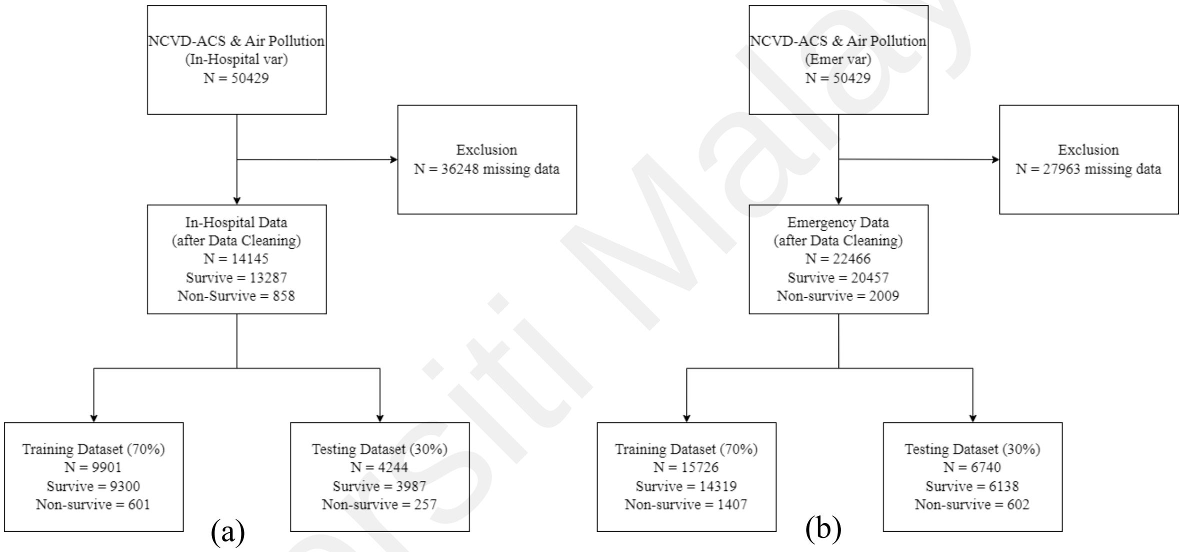


Figure 3.4: The flowchart indicating the raw number of instances before and after data cleansing in NCVD-ACS and air pollution data for (a) In-Hospital Variables (b) Emergency Variables.

Besides, the allocation of 70% - 80 % of original data for training and 30% - 20% for testing gives the best outcomes for several empirical results. In Figure 3.5 below is the suggested optimal ration of the train-test split according to the size of the dataset, as our dataset ranges at ~10,000 cases or more, in our study, data is split into 70:30 manner (Gholamy, et al., 2018).

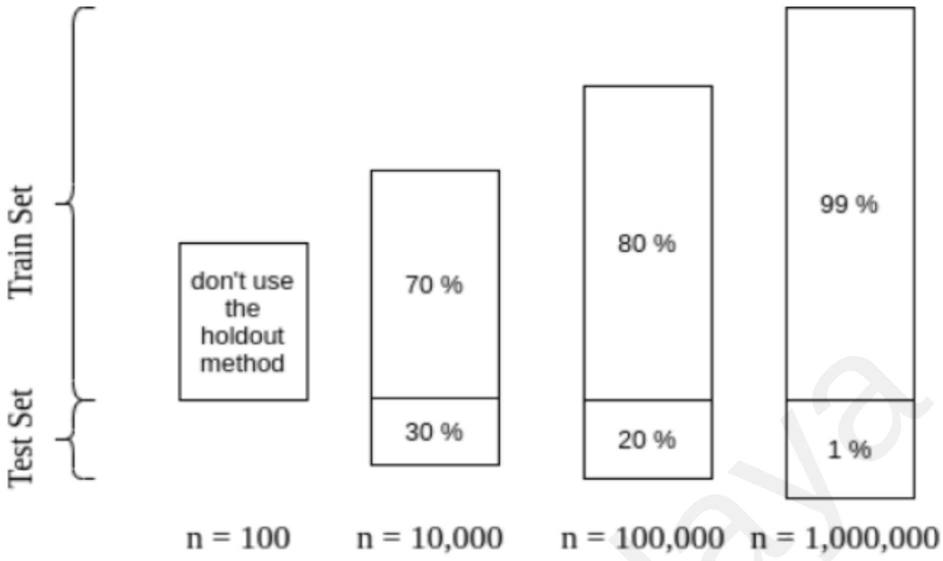


Figure 3.5: The suggested optimal ratio of the train-test split according to the size of the dataset (Photo source from Gholamy, et. al., 2018).

Instead of dividing the data into training, validation, and test subsets and conducting holdout cross-validation on the validation set. The data was partitioned into training and testing sets, and k-fold cross-validation was performed. This method is considered more effective than the traditional train-validation-test split cross-validation (Hsieh, et al., 2019). In K-fold cross-validation, the input data is divided into 'k' number of folds, such as k=5, where the dataset will be split into 5 folds and the model will be iterated, trained and evaluated for 5 times, with each fold used once for testing and the remaining folds used for training (Ajitesh, 2023). It is used to validate the performance of the developed ML models to ensure the best model is selected (Figure 3.6) (Rukshan, 2020).

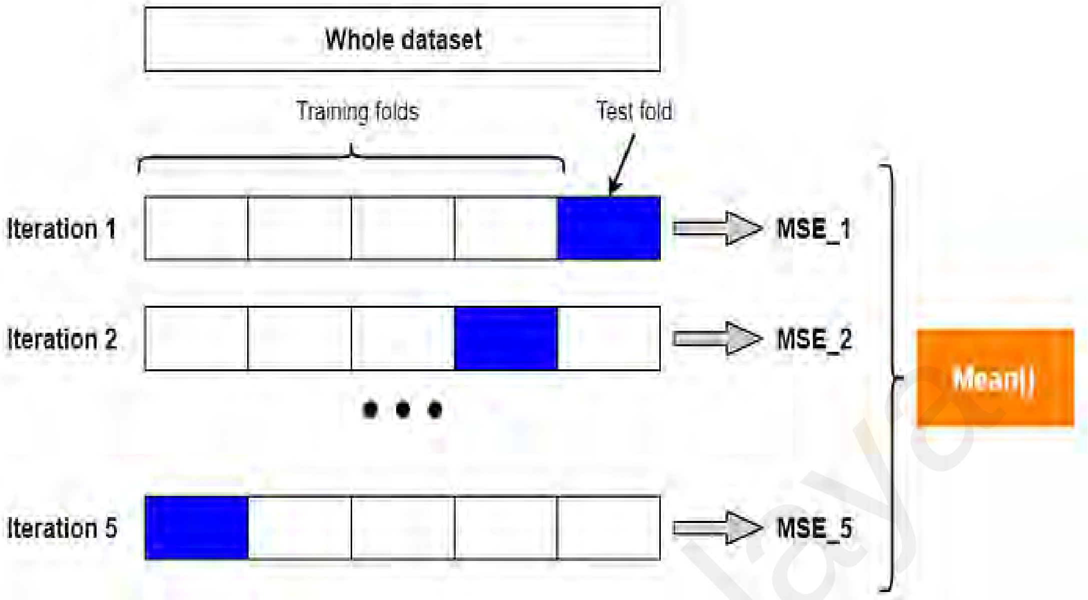


Figure 3.6: An illustration of 5-fold cross-validation for evaluating machine learning (ML) model's performance (Photo sourced from Rukshan, 2020).

5-folds cross validation is applied in this study, the lower number of folds may affect the model suffer from high bias, where only a smaller portion of the dataset are trained. On the other hand, higher number of folds could lead to higher variance, where there is possibility training the entire dataset may result in overfitting. Hence, cross-validation with 5 folds provides a decent balance between bias and variance, ensuring that the model performs well with unseen data (Kohavi, 1995). Furthermore, applying 5-folds cross-validation is computationally less expensive in compared to higher number folds, result in faster and better performance during model development (John Lu, 2010).

3.7 Data Balancing

Data imbalance can be defined as discrepancy in the number of instances for each class within a dataset, which causes classifiers performance to deteriorate as the model are not able to learn the features of the less represented class, often found in medical domain (Domingues, et al., 2018). Data balancing is important in ML development, particularly in classification tasks, where the outcome of the study is binary. The input data must be balanced to ensure the accuracy and the performance of the developed ML model. An imbalanced dataset may lead to biased predictions and affect the accuracy and the model performance, as the developed model becomes biased towards the majority class (Batista, et al., 2004).

In this study, data balancing is applied in the classification model that predicts the probability of ACS mortality in ACS patients in presence of air pollution. In the in-hospital features dataset consists of 14,145 instances, and in the emergency selected features consists of 22,466 instances, with two classes: ‘survive’ and ‘non-survive’. The distribution of the target features we observed in the selected feature dataset, ‘survive’ patients has 9,300 instances, while ‘non-survive’ patients have only 601 instances shown in Figure 4(a). For the emergency dataset, ‘survive’ patients consist of 14,319 instances, while ‘non-survive’ patients only consist of 1,407 instances shown in Figure 4(b). The number of ACS’s patients that are labelled as ‘survive’ is higher than the ‘non-survive’ patients, and this might cause significant imbalanced in the dataset and may lead to biased when making predictions in our classification model.

There are several data balancing approaches to improve the ML classifier performance:

1. Oversampling: Suitable for small dataset, by increasing the number of minority class samples by duplication or resampling randomly (Domingues, et al., 2018).
2. Under sampling: Suitable for big dataset, by deleting the majority class instances (Jadhav, et al., 2022).
3. Synthetic Data Generation: New instances are generated based on samples of the minority classes, such as by using the ROSE algorithm (random over-sampling examples)

The ROSE method is used to perform data balancing in our study, ROSE package contains functions for handling binary classification problems with imbalanced classes. A smoothed bootstrap method is used to make artificially balanced samples that can help with both the estimation and accuracy evaluation parts of a binary classifier when there is a rare class. This package has well-defined accuracy functions for quickly completing tasks. ROSE can maintain the overall data structure while generating synthetic samples for both classes. It can also deal with both continuous and categorical data (Lunardon, et al., 2014).

Oversampling and under sampling approaches in not chosen our study due to their inherent drawbacks. The oversampling method can lead to overfitting of the ML model as it duplicates information from the minority class. Conversely, the under-sampling method removes many instances randomly until the dataset is balanced, which could result in the loss of potentially useful and important information. In cases where the data is heavily imbalanced, as in this study, these two approaches are deemed unsuitable, since it could negatively impact the performance of our classification model (Jadhav, et al., 2022).

3.8 Data Normalization

In ML, data normalization is an important data pre-processing step that converts continuous variables in a dataset into a common scale to ease comparison and analysis. Normalization can make classification and grouping more accurate by eliminating of various scales and units of measurement in the original data (Starovoitov & Golub, 2021).

Min-max normalization is a linear scaling technique that scales feature values to the range of [0, 1], where the minimum and maximum value of a feature will normalize to be in the range 0 to 1 (Serafeim, 2020). The formula for min-max normalization is as follows:

$$x_{scaled} = \frac{x - \min(x)}{\max(x) - \min(x)} \qquad (3,1)$$

Data normalization was applied to continuous variables, including age, heart rate, high-density lipoprotein, low-density lipoprotein, fasting blood glucose, NOx, SO_2, O_3, and PM10, using the min-max normalization approach. Research has demonstrated that data normalization can significantly enhance the accuracy and efficiency of ML algorithms (Tina & Sherekar, 2013; Jiawei, et al., 2012; Pedregosa, et al., 2011). Given these findings, the min-max normalization technique was considered a suitable choice for our study.

3.9 Baseline Characteristics

Baseline characteristics are the descriptive information collected at the initial stage of a study about the participants. Baseline data include medical, demographic, and other types of information (Norwegian Research Council, 2017). It has the potential to increase statistical efficiency by improving the ability to derive meaningful conclusions from given data (Holmberg & Andersen, 2022).

Statistical Package for the Social Sciences (SPSS) version 26.0 was utilized to conduct all analyses of baseline characteristics for the four distinct models. Continuous variables are presented as mean ± standard deviation, whereas categorical variables are shown as frequency and percentage.

In this study, the baseline characteristics in the regression model were analyzed to estimate the ACS patients' hospitalization and mortality rate in relation to air pollution. Multiple linear regression was employed to identify significant variables, given the continuous nature of the predictors.

In the classification model, a univariate analysis was conducted to identify significant factors. Our dataset includes both continuous and categorical inputs; as such, the Chi-square test was applied for categorical variables like Killip class, statin, and others. The normality of continuous data variables such as patients' age, heart rate, HDLC, LDLC, FBG and air pollutants was tested using the Kolmogorov-Smirnov (K-S) Test ($p<0.05$). Based on these normality tests, the data are not normally distributed, the Mann-Whitney test was applied to these continuous variables that were not normally distributed. A p-value of <0.05 from the K-S test indicates a deviation from normality. Histograms and test results are presented in appendix D. This approach ensures a clear structured analysis of the data in our study.

3.10 Machine Learning (ML) Model Development

Two types of ML model are developed, regression and classification. Algorithms that are used to develop the models are those commonly cited in literature; the summary of the ML models used is as shown in Table 3.7:

(a) Regression Models: The regression algorithms are used to predict ACS's hospitalization and mortality rate in the presence of air pollution. The five regression algorithms are: linear regression, SVM, RF, XGBoost and ensemble learning. The naïve bayes algorithm is not used for the regression model, as naïve bayes is only used in classification setting (Leung, 2007).

(b) Classification Models: The classification algorithms are developed to predict the post ACS mortality risk for in-hospital and emergency patients with daily air pollution readings. Classification models predict categorical or binary outcomes, the models are trained based on selected features and classify ACS's patients into two categories: "Survive" or "Non-survive". The ML classification algorithms developed are logistic regression, naïve bayes, SVM, RF, XGBoost and ensemble learning.

Table 3.7: Summary of objectives, expected outcomes, and ML type and algorithms utilized in this Study.

Objectives	Expected Outcome	Machine Learning Type	Machine Learning Algorithm
Predict ACS hospitalization rates	Number of ACS admission cases per day	Regression	1. Linear Regression 2. Support Vector Regression 3. Random Forest 4. Extreme Gradient Boosting 5. Ensemble Learning - GLM
Predict ACS mortality rates	Number of ACS mortality cases per day	Regression	1. Linear Regression 2. Support Vector Machines 3. Random Forest 4. Extreme Gradient Boosting 5. Ensemble Learning - GLM
Predict probability of mortality for ACS patients based on selected variables	Probability of mortality of ACS's patient.	Classification	1. Logistic Regression 2. Support Vector Machines 3. Naïve Bayes 4. Random Forest 5. Extreme Gradient Boosting 6. Ensemble Learning - GLM
Predict probability of mortality for ACS patients based on emergency variables	Probability of mortality of ACS's patient.	Classification	1. Logistic Regression 2. Support Vector Regression 3. Naïve Bayes 4. Random Forest 5. Extreme Gradient Boosting 6. Ensemble Learning - GLM

The ML model development flowchart is shown in Figure 3.7 below. The process begins with data preparation, that involves cleaning, partitioning, balancing and normalization, then proceeds to model development, model hyper parameter tuning.

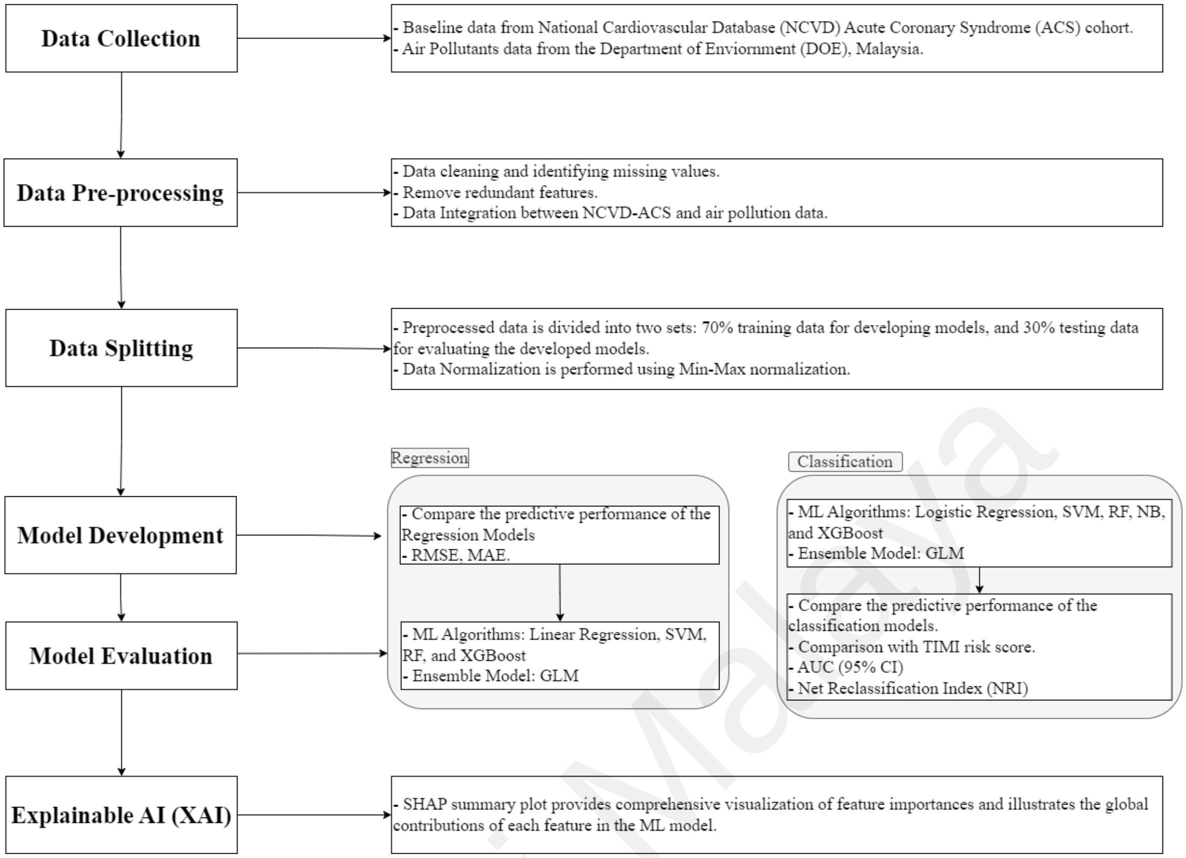

Data Collection
- Baseline data from National Cardiovascular Database (NCVD) Acute Coronary Syndrome (ACS) cohort.
- Air Pollutants data from the Department of Enviornment (DOE), Malaysia.
Data Pre-processing
- Data cleaning and identifying missing values.
- Remove redundant features.
- Data Integration between NCVD-ACS and air pollution data.
Data Splitting
- Preprocessed data is divided into two sets: 70% training data for developing models, and 30% testing data for evaluating the developed models.
- Data Normalization is performed using Min-Max normalization.
Regression
Classification
Model Development
- Compare the predictive performance of the Regression Models
- RMSE, MAE.
- ML Algorithms: Logistic Regression, SVM, RF, NB, and XGBoost
- Ensemble Model: GLM
Model Evaluation
- ML Algorithms: Linear Regression, SVM, RF, and XGBoost
- Ensemble Model: GLM
- Compare the predictive performance of the classification models.
- Comparison with TIMI risk score.
- AUC (95% CI)
- Net Reclassification Index (NRI)
Explainable AI (XAI)
- SHAP summary plot provides comprehensive visualization of feature importances and illustrates the global contributions of each feature in the ML model.

Figure 3.7: Machine learning (ML) development workflow.

Figure 3.8 illustrates the detailed flow for regression model development. Figure 3.9 shows the detailed flow for classification model development. These figures detail the sequence of steps in our ML application, including model development, hyper parameter tuning, and the selection of the best-performing model.

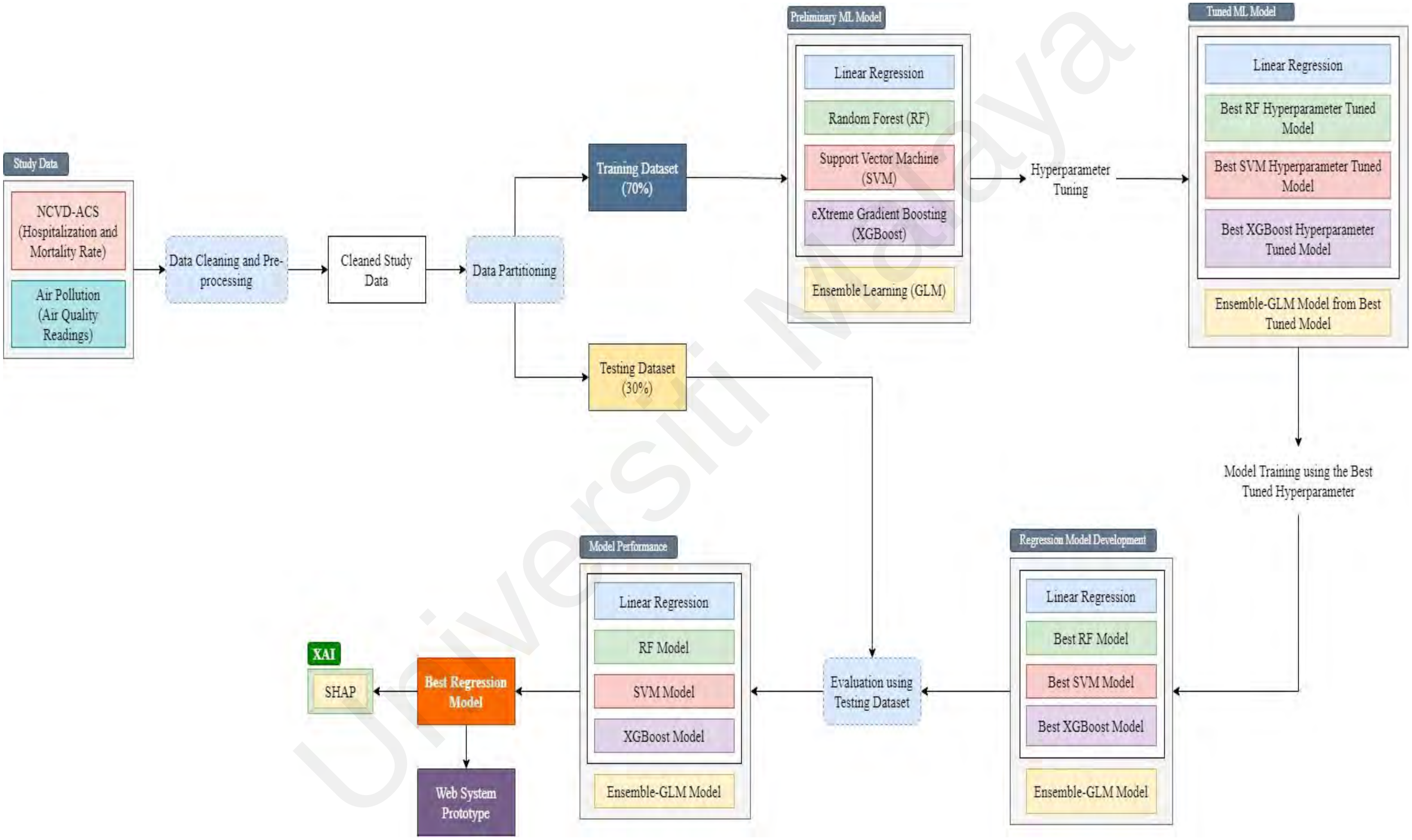

Study Data
NCVD-ACS (Hospitalization and Mortality Rate)
Air Pollution (Air Quality Readings)
Data Cleaning and Pre-processing
Cleaned Study Data
Data Partitioning
Training Dataset (70%)
Testing Dataset (30%)
Preliminary ML Model
Linear Regression
Random Forest (RF)
Support Vector Machine (SVM)
eXtreme Gradient Boosting (XGBoost)
Ensemble Learning (GLM)
Hyperparameter Tuning
Tuned ML Model
Linear Regression
Best RF Hyperparameter Tuned Model
Best SVM Hyperparameter Tuned Model
Best XGBoost Hyperparameter Tuned Model
Ensemble-GLM Model from Best Tuned Model
Model Training using the Best Tuned Hyperparameter
Regression Model Development
Linear Regression
Best RF Model
Best SVM Model
Best XGBoost Model
Ensemble-GLM Model
Evaluation using Testing Dataset
Model Performance
Linear Regression
RF Model
SVM Model
XGBoost Model
Ensemble-GLM Model
Best Regression Model
XAI
SHAP
Web System Prototype

Figure 3.8: The flowchart of the regression ML predictive models' development.

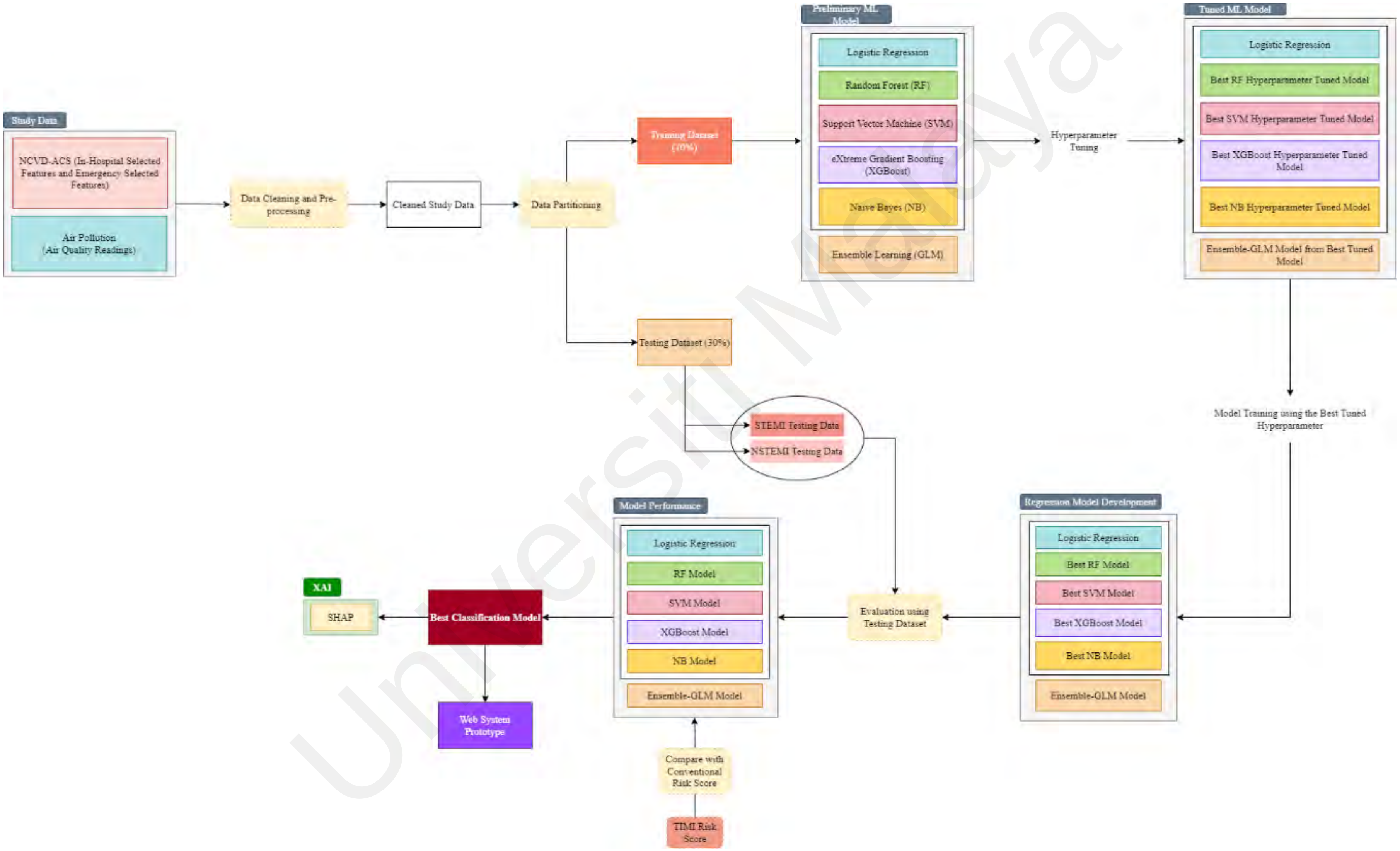


Figure 3.9: The flowchart of the classification ML predictive models' development.

3.10.1 Algorithm Overview

The regression models in this study utilize linear regression instead of logistic, aimed to predict the ACS's hospitalization and mortality rate. Conversely, the classification models employed logistic regression to replace linear regression, focusing on classifying outcomes into categories. Despite the difference in outcomes, continuous for regression and categorical for classification, the structure and development process of both types of models were largely parallel.

The selection of ML algorithms for both model types was determined by their diverse underlying methodologies, which allowed for the exploration of a variety of approaches in capturing relationships within the data. Additionally, Generalized Linear Model (GLM) was employed as a meta-learner in the EL model. This approach was consistent across both the regression and classification tasks, further highlighting the similarities in our model development process.

(a) Linear Regression

The linear regression model was fitted using *lm()* function to predict hospitalization rates and number of mortalities of ACS patients in the presence of air pollution. The *lm()* function lacks tuning option, thus the default value of standard linear model directly with the dataset. Despite the lack of hyper parameters in linear regression, the model was enhance through resampling with the *trainControl()* function through 5-fold cross-validation to increase the robustness of our model. This process splits the data into five subsets, validating the model systematically, thereby achieving a reliable performance of the model across varied data samples. Linear regression is only applied in the development of the regression model.

(b) Logistic Regression

The logistic regression is built using the *glm()* function to fit the classification model. (Kassambara, 2018) stated that there are no tuning hyper parameters in logistic regression. Hence, the default parameter was used to conduct binary classifications on the selected variables and emergency variables for the ACS dataset in the presence of air pollution. Similar to linear regression model, we enhance the model performance through resampling method. The *tunelength* function was set 10, where the longer the tune length allows the algorithm to examine more potential models and possibly find a better one, however, this process is computationally expensive. 5-fold cross-validation was also carried out in the model building.

(c) Support Vector Machine (SVM)

The SVM (Cortes & Vapnik, 1995) model is used for both regression and classification. SVM algorithm is capable of handling linear and non-linear relationships between the predictor and target variables (Rana, 2015).

'*svmLinear'* was used to develop the model. SVM can utilize different types of kernels, including linear and radial basis function (RBF) kernels. The linear kernel corresponds to linear mapping, while RBF kernel is commonly used to handle non-linear relationships.

The hyper parameter for SVM is Cost, *C*, which determines the possibilities 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. Optimum *C* value can be chosen via the highest ROC value.

(d) Random Forest (RF)

RF (Breiman, 2001) algorithm is used for both regression and classification. It is one of the ensemble learning methods that operates by constructing multiple decision trees (Fawagreh, et al., 2015). The RF approach is a regression tree method that enhances prediction accuracy by aggregating bootstrap data and randomizing predictors (Rigatti, 2017).

There are two tuning parameters in RF: *ntree* and *mtry*. '*ntree'* is the number of trees to grow, which must be large enough to provide OOB error stabilization. The default value is 500 in the caret package, the larger number of trees can lead to a more robust model. However, after a certain number of trees, the model's predictive performance may not significantly improve since the additional trees tends to be highly correlated with predictions made by the existing trees (Breiman, 2001).

'*mtry'* is the maximum number of features considered for splitting a node, which can range from 1 to the total number of variables. The default value of *mtry* is √N, where N is the total number of variables. In general, using the default values for *ntree* and *mtry* can yield good results.

During the model development process, the model is trained with varying numbers of trees (*ntree*), specifically 500, 1000 and 1500. To determine the optimal number of variables to be considered at each split (*mtry*), grid search with a tuneLength of 10, combined with 5-fold cross-validation was applied.

(e) Naïve Bayes (NB)

NB (Bayes, 1968) algorithm is only applied in classification model for this study. It is a probabilistic algorithm based on applying Bayes' theorem with strong independence assumptions, commonly used for classification tasks.

To implement the NB classification model, the '*nb*' function from the 'caret' package in R was used. This allows the function to begin training the model and perform prediction tasks. The hyper parameters of the NB algorithm are '*fit_prior*', were tuned to optimize the model's performance. '*Fit_prior*' was modified to determine if the model should learn the prior probabilities from the training data or presume uniform prior probabilities (Chong & Shah, 2022).

(f) Extreme Gradient Boosting (XGBoost)

XGBoost (Chen, et al., 2017) is used for both regression and classification model development in this study. It is a widely used ensemble learning technique that utilizes gradient boosting decision trees to produce reliable predicted performance. '*xgbTree*' function is used to build the model from the caret package in R. Unlike other ML algorithms, XGBoost offers a larger set of hyper parameters to tune, providing an opportunity for fine-tuned model optimization.

The specific hyper parameters we tuned in this study include 'max_depth', '*min_child_weight*', '*gamma*', '*subsample*', and '*colsample_bytree*'. Each hyperparameter plays a distinct role in the model.

- '*max_depth*': maximum depth of the individual regression estimators.
- '*min_child_weight*': a regularization parameter that helps control overfitting.
- '*gamma*': specify minimum loss reduction required to make further partition on a leaf node of the tree.
- '*sub_sample*': the fraction of observations to be randomly sampled for each tree, introducing randomness into the model building process.
- '*colsample_bytree*': the fraction of columns to be randomly sampled for each tree to prevent overfitting as well.

The tuning of these parameters is an iterative process, guided by cross-validation to ensure that the selected values were relevant to unobserved data. Tuning these parameters can help prevent overfitting and improve the accuracy of the model (Saraswat, 2016).

(g) Ensemble learning (EL)

EL was developed for both regression and classification as well to improve the strengths of multiple models, thereby improving predictive accuracy and reliability. This approach combines the predictions from several base learners to generate a final output. The base learners for regression model used are linear regression, SVM, RF, and XGBoost. For classification model, the base learner used are logistic regression, NB, SVM, RF, and XGBoost.

In ensemble learning, the 'meta-learner' is used to combine the predictions from the base learners. The Generalized Linear Model (GLM) is selected as meta-learner for both models. The choice of GLM was based on its flexibility and its ability to handle various types of distributions, linear relationships, and non-linearity compared to other meta-learners such as RF and Gradient Boosting Methods (GBM) (Song, 2013).

The process of creating an ensemble model involved training the base learners on the dataset, each producing a set of predictions. These predictions were then used as input features for the meta-learner, the GLM, which was trained to make the final prediction. The library '*caretEnsemble*' is used to ensemble the base models, using the '*glm*' as our method in building the ensemble model.

To ensure optimal performance, the hyper parameters of the GLM meta-learner were also tuned using a grid search cross-validation approach. The ensemble model was then evaluated and validated using the same performance metrics and procedures applied to the base learners.

3.10.2 Machine Learning Model Hyperparameter Tuning

Hyperparameter tuning is a crucial step in the ML process. It is selecting a set of optimal hyperparameters for a learning algorithm. It has the potential to improve a model's performance by getting the right combination of hyperparameters, different ML models require the tuning of different hyperparameters. If these are not explicitly defined, the algorithm defaults to pre-set values. This approach ensured that the ML models were robust and accurately represented the relationship between air pollution and ACS incidence in our study.

The '*caret*' package in R is applied for the development of both ML models for regression and classification. The package contains functions to streamline the model training process for complex regression and classification problems (Kuhn & contributors, 2023). Caret package is used instead of other library and packages offer in R, is to ensure the consistent outcomes regardless of the model complexity and ease our analysis. The functions used are:

(a) *train_control(method = , number = , search = , savePredictions = TRUE , classProbs = TRUE, summaryFunction = twoClassSunmmary, allowParallel = TRUE)*: To specify the resampling scheme and used to set parameters and control the training process.

Method = "cv": specified cross-validation resampling method used.

Number = 5: indicates the number of resampling iterations which is the number of folds in k-fold cross-validation. In our study, 5-fold cross-validation is applied.

Search = "grid": defines the type of grid search performs when tuning model parameters.

classProbs = "TRUE": The class probabilities for each prediction will be saved, this is only applicable in classification models.

summaryFunction = “twoClassSummary”: It computes class probabilities in addition to the class predictions, this is only applicable in classification models.

allowParallel = TRUE: It allows for parallel processing, which speeds up the computation.

(b) *train*(*x, y, method=, trControl = train_control, metric= , tuneGrid =*): Act as the workhorse of *caret*, handling several parameters crucial to our model development.

x, y: are the features and the target variables.

Method: the modelling method is defined, such as “rf”, “glm”, “xgbTree”.

trControl = train_control(): is where ‘trControl()’ object is passed, which specifies the resampling method.

metric: It is used to optimize the model. “RMSE” is used for regression models, “AUC” is used in classification models.

tuneGrid: used to specify the hyperparameter grid to search over, the hyper parameters tuned is presented in table 7.

The application of these functions allowed us to fine-tune models and select the best model by tuning the algorithms hyperparameters. Cross-validation was used to avoid overfitting and increase generalizability of the developed models. In this study, 5-fold cross-validation (*k=5*) was applied.

The grid search was used to find the best model by choosing the tuning of the values of the parameters. The parameters for each model were chosen based on recommendations from the literature and the default settings in the respective R packages.

Table 3.8 presents the summary of the main hyperparameters utilized in our study. The optimized parameters for each regression and classification ML models are shown in Table 3.9 and Table 3.10 respectively.

Table 3.8: Summary of the main hyperparameters utilized in this study.

Algorithm (R packages)	Key Hyperparameters	Suitable Task
Linear Regression (*lm*)	None	Regression
Logistic Regression (*glm*)	None	Classification
SVM (*svmLinear*)	cost (Penalty parameter C of the error term), gamma (Kernel coefficient for 'rbf', 'poly' and 'sigmoid'	Classification, Regression
Random Forest (*randomForest*)	ntree (Number of trees to grow), mtry (Number of variables randomly sampled as candidates at each split)	Classification, Regression
Naive Bayes (*nb*)	Fit_prior (Fit Prior), usekernel (Whether kernel density estimates should be computed for numeric attributes)	Classification, Regression
XGBoost (*xgboost*)	eta (Learning rate), max_depth (Max depth per tree), min_child_weight (Minimum sum of instance weight), subsample (Subsample ratio of the training instances), colsample_bytree (Subsample ratio of columns when constructing each tree), nrounds (Number of boosting rounds), gamma (Minimum loss reduction), alpha (L1 regularization), lambda (L2 regularization)	Classification, Regression
GLM Ensemble (*glmnet*)	alpha (Elastic net mixing parameter), lambda (Regularization parameter)	Classification, Regression

Table 3.9: The hyperparameters values for optimum ML model performance for regression models.

Regression Machine Learning (ML)	
ACS Hospitalization Rate Model	
ML Algorithm	**Parameters**
Linear Regression	Default

Random Forest (RF)	Mtry = 1
	Ntree = 1000

Support Vector Machine (SVM)	Kernal = Linear
	C = 3

XGBoost	max_depth = 8
	min_child_weight = 12
	nrounds = 100
	eta = 0.3
	subsample = 0.6

Table 3.9, continued.

ML Algorithm	Parameters
Ensemble learning (EL)	Base learner: linear regression, RF, SVM, NB, XGBoost Meta learner: glm
ACS Mortality Rate Model	
ML Algorithm	**Parameters**
Linear Regression	Default

Random Forest (RF)	Mtry = 2
	Ntree = 1000

Support Vector Machine (SVM)	Kernal = Linear
	C = 5

XGBoost	max_depth = 5
	min_child_weight = 8
	nrounds = 300
	eta = 0.2
	subsample = 0.6

Ensemble learning (EL)	Base learner: linear regression, RF, SVM, NB, XGBoost Meta learner: glm

Table 3.10: The hyperparameters values for optimum ML model performance for classification models.

Classification ML	
In-Hospital Model	
ML Algorithm	**Parameters**
Logistic Regression	Default

Random Forest (RF)	Mtry = 2
	Ntree = 1000

Support Vector Machine (SVM)	Kernal = Linear
	C = 3

Naïve Bayes (NB)	Fit_prior = TRUE
	useKernal = TRUE

XGBoost	max_depth = 7
	min_child_weight = 3
	nrounds = 300
	eta = 0.02
	subsample = 0.7

Ensemble learning	Base learner: logistic regression, RF, SVM, NB, XGBoost Meta learner: glm

Table 3.10, continued.

Emergency Model	
ML Algorithm	**Parameters**
Logistic Regression	Default

Random Forest (RF)	Mtry = 2
	Ntree = 1000

Support Vector Machine (SVM)	Kernal = Linear
	C = 3

Naïve Bayes (NB)	Fit_prior = TRUE
	useKernal = TRUE

XGBoost	max_depth = 7
	min_child_weight = 3
	nrounds = 300
	eta = 0.02
	subsample = 0.7

Ensemble learning (EL)	Base learner: logistic regression, RF, SVM, NB, XGBoost Meta learner: glm

3.11 Model Evaluation

These metrics provide a comprehensive evaluation of the performance of the ML models in both regression and classification tasks.

3.11.1 Regression Algorithm Performance Metrics

Evaluating the performance of regression models primarily involves the assessment of the predicted values' closeness to the actual values. Two metrics were used to evaluate the performance of the regression models: Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE).

1. **Root Mean Squared Error (RMSE):** RMSE is a commonly used evaluation metric for regression models. It measures the square root of the average of the squared differences between the predicted and actual values. It is useful in determining how

far off the predictions are from the actual values. Hence, the lower the RMSE indicates the better performance of the model.

2. **Mean Absolute Error (MAE):** MAE measures the average magnitude of the errors in a set of predictions. The difference between MAE and RMSE, MAE uses the absolute value of error, rather than the squared error. Thus, MAE is suitable for data that is not normally distributed and when there are outliers in the data. Like RMSE, a lower MAE indicates a better predictive performance of the regression model.

While both RMSE and MAE are the evaluation metrics for regression model performance. According to Hyndman & Koehler (2006), RMSE is generally preferred over MAE because it emphasizes significant errors, which can be more significant in forecasting scenarios. Furthermore, RMSE is commonly used for evaluating regression models due to its ability to penalize large errors more significantly (Palacio-Niño & Berzal, 2019).

In this study, both RMSE and MAE are used to gain insights into the model's overall performance, providing valuable insights into their predictive capabilities and identifying any potential areas for improvement.

3.11.2 Evaluation of Classification Models

For classification models, the model's ability to correctly classify the target variable was assessed using the following metrics: Accuracy, Precision, Recall, F1-Score, and the Area Under the Receiver Operating Characteristic Curve (AUROC).

In this study, the classification model is used to predict the mortality of ACS patients in the presence of air pollution. The model's features are based on previous research (Kasim et al., 2022) and emergency features identified by cardiologists, which are combined with air pollution variables. These features are used to estimate the mortality of ACS patients in the presence of air pollution.

The commonly used evaluation for classification model is the AUROC. According to Huang & Ling (2005) and Fawcett (2006), the AUROC has been utilized in medical diagnosis since 1970s, instead of accuracy, AUROC should be used to and compare classifiers because AUROC is a more accurate measure in general, and simple classification accuracy is usually a poor statistic for evaluating performance. This is because AUROC 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, recall, and F1 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.

Furthermore, AUROC was employed as an indicator in many of the medical diagnoses to assess the performance of the model they constructed. In Suzuki, et al. (2019) study, AUROC is used as comparative performance of ML models for predicting early mortality in acute heart failure, logistic regression performed better than other ML models with the AUC of 0.794. Similarly, the AUROC is employed to discuss and compare the result of several supervised ML algorithms for predicting the risk of coronary heart disease (Beunza, et al., 2019). AUROC is used as an evaluation metric for ML models because it is a measure of the probability that a model will correctly classify a positive instance as positive. AUROC is not affected by the prevalence of the positive class, which makes it a more robust metric than accuracy. All in all, these studies chose AUROC due to it is unaffected by class imbalances.

Apart from AUROC, confusion matrix is one of the performance measurements for ML classification, where it is a table with 4 different combinations of predicted and actual values, the output can be binary or more classes. In R, the library ‘ggplot’ was used to construct the confusion matrix table. In this study, the true positive class of our outcome was set to be 0 (died). The four main values in a confusion matrix are:

i) True Positives (TP): The model accurately classifies the patients as non-survivors.

ii) False Positives (FP): The model incorrectly classifies survival patients as non-survivors.

iii) True Negatives (TN): The model accurately classifies that the patients are survivors.

iv) False Negatives (FN): The model incorrectly classifies non-survivor patients as survivors.

Table 3.11: Confusion matrix for classifying the ACS patients' outcome.

		Actual Outcome	
		0 (Dead)	1 (Alive)
Predicted Outcome	0 (Dead)	TP	FP
	1 (Alive)	FN	TN

False Negatives (FNs) must take priority over False Positives (FPs) in our predictive model due to the critical nature of the medical domain. This is because the cost of incorrectly classifying an ill patient as healthy (FN) can result in a failure to provide essential medical treatment, which is significantly greater than the cost of incorrectly classifying a healthy person as sick (FP).

In addition to the AUROC and the confusion matrix, several other performance metrics were utilized to evaluate the performance of the models within the context of this study.

1. Accuracy: Measures the proportion of true results (both true positives and true negatives) in the population.
2. Precision (Specificity): Precision evaluates the proportion of true positives out of the predicted positives. It is also known as true positive rate.
3. Recall (Sensitivity): Recall measures the proportion of actual positives that are correctly identified. It is known as true negative rate.
4. F1-Score: The F1-score is the harmonic mean of precision and recall, providing a balanced measure when class distribution is uneven.

3.12 Explainable Artificial Intelligence (XAI)

Due to the black box nature of ML algorithms, there was a lack of understanding when using ML in the medical domain. This makes it difficult to understand how ML models make decisions, which can be problematic in critical areas such as medical decision-making. The release of explainable Artificial Intelligence (XAI) offers a solution in providing transparency and explaining the complex ML models on how it makes their decisions. Implementing XAI in the medical domain is critical because it provides transparency to healthcare professionals, allowing them to understand and trust the ML models' predictions (Rao, et al., 2022; Zeng, 2022). In this study, SHapley Additive exPlanations (SHAP) is used to interpret the ML model prediction.

The 'shap' library is used for computing SHAP values. The 'shap' library offers a unified measure of feature importance and effects. The code implementation involved first training our selected ML models and making predictions. Then, the SHAP explainer was fitted on the trained model and computed the SHAP values for the predictions.

SHAP values provide a fair allocation of each feature's contribution to each prediction in comparison to the baseline prediction. It offers both global interpretability (overall importance of features).

The use of XAI improved the interpretability and trustworthiness of the ML models, increasing their potential applicability in the medical field, particularly in predicting ACS hospitalization rates, ACS mortality rates, and the probability of mortality for ACS patients. SHAP was used to analyze the models that performed the best based on the evaluation metrics. Chapter 4 presents the findings and discussions from these analyses.

3.13 Comparison of ML Classification Models with the TIMI Risk Score for Predictive Validation

In this study, the ML classification models were compared to the TIMI risk score to validate the ML predictive abilities. This comparative study was designed to demonstrate the advantage of the ML models in predicting ACS patient mortality risk in the presence of air pollution. The 30% from the original dataset is reserved for testing and comparison with the TIMI score, the testing dataset is further divided into two distinct categories: ST-elevation myocardial infarction (STEMI) and non-ST-elevation myocardial infarction (NSTEMI) to align the evaluation with the clinical application of the TIMI Score. This facilitated alignment with the patient risk category cutoff points specific to TIMI clinical practice.

The ROC curves for each ML model as well as the TIMI risk score are derived to compare the performance specifically for patients with STEMI and NSTEMI. Additionally, graphs on the mortality rate in relation to the TIMI risk score and the best performing ML models' percentile values were derived to differentiate between the high- and low risk patients based on clinical practice and existing literature (Correia, et al., 2014). A high risk of mortality was defined as a probability risk of mortality of more than 8% like reported by Correia et al.

(2014). To further determine the statistical significance of the trend analysis, the rate of mortality graphs was also then tested for the trend in terms of a p-value.

3.14 Net Reclassification Improvement Index (NRI) for Classification Model

Net Reclassification Improvement (NRI) is a statistical measure used to evaluate the performance of predictive models. It measures the improvement in classification of individuals into higher or lower risk categories when a new model is compared to an existing risk strategy (Zhou, et al., 2022). The NRI facilitates a comparative analysis of the best ML classification models and the conventional TIMI risk score in predicting the risk associated with ACS in the presence of air pollution.

Correia et al. (2014) discovered a cut-off value between low and high-risk patients on mortality for the best ML models based on the percentage of mortality using NRI. Morrow et al. (2000) and Antman et al. (2008) define appropriate cut-off points for STEMI and NSTEMI/UA patients, respectively, for the TIMI risk score.

NRI quantifies how well a new mortality risk assessment approach inspires appropriate categorization between categories. It essentially assesses the net improvement in patient classification by employing an unconventional approach (Pencina, et al., 2008). NRI was used to distinguish between the traditional TIMI risk score and ML classification algorithms in terms of discrimination power.

NRI employs reclassification tables to examine the additive benefit derived from reclassifying patients using a different mortality assessment methodology. Given that the study targets a binary variable, the two-category NRI was used to assess the efficacy of the best models.

An NRI of zero signifies equal discriminatory abilities between the new and old models. A negative NRI suggests that the new model failed to discriminate as well as the old model between high-risk and low-risk categories, while a positive NRI indicates superior discrimination by the ML model. The NRI scale extends from -2 to 2, with 2 signifying perfect discrimination by the new model and complete failure by the old, and -2 indicating the reverse. Finally, the ANOVA test was used to calculate the p-value, comparing the probability of mortality predictions, and identifying its significance from our ML models with the established TIMI risk score.

3.15 Web System Analysis and Design

The web system integrates the best ML models for regression and classification. This section covers web system design and development, including requirement analysis, system architecture, user interface design, and website wireframes. To evaluate the usability of the web system, system usability testing was carried out. This web system's primary goal is to visualize and predict ACS events related to air pollution.

3.15.1 Prototyping Model

The prototyping model is used to develop the web system. It is an iterative process where the web prototype is built, tested, and refine until the user is satisfied with the design (Martin, 2023). The prototype is iteratively refined with user involvement. The final prototype was converted into the web system. Below are the phases of the prototyping model and the flowchart is shown in Figure 3.10.

1. Requirement Analysis

   The prototyping model starts with requirement analysis. In this phase, the requirements of the system are gathered and stated clearly. During this phase, the users of the system are interviewed to understand the expected outcomes of the system.

2. Design

   After understanding the user requirements, the second phase is considered as a preliminary design of the system, which gives a brief description of the system created based on the user's request. In this phase, the basic layout of the website and architecture is designed, the wireframes are created to help in visualizing the design.

3. Prototyping

   The core features are implemented during this phase. The basic functionality of the system is created, including integration of ML models, geospatial mapping, and basic mortality calculators.

4. User Evaluation

   The prototype is then presented to users for evaluation. During this phase, the users are encouraged to interact with the system and provide feedback on the design, functionality, and overall experience. The feedback and suggestions are gathered for further discussion at the later phase.

5. Refining Prototype

   The prototype is further refined and improved according to the user's feedback and suggestions.

Step 2 to step 5 is iterated as necessary until the end users are satisfied with the design prototype.

6. User Approval

   The final prototype is deployed based on the approval by the users once all the requirements set are met.

7. Deployment

   With the approved final prototype, moving on into the deployment phase. During this phase, the approved system prototype is deployed on a live server, with all necessary integrations and setups.

8. Testing

   Testing is the last phase of the prototyping modeling process, the final web system is evaluated and tested. The usability of the system and user experience is tested using the system usability scale (SUS). Final improvement of the system is carried out based on the comments from the SUS questionnaire.

9. Release

   After completing the testing phase and addressing the remaining issues, the web system is released for the users.

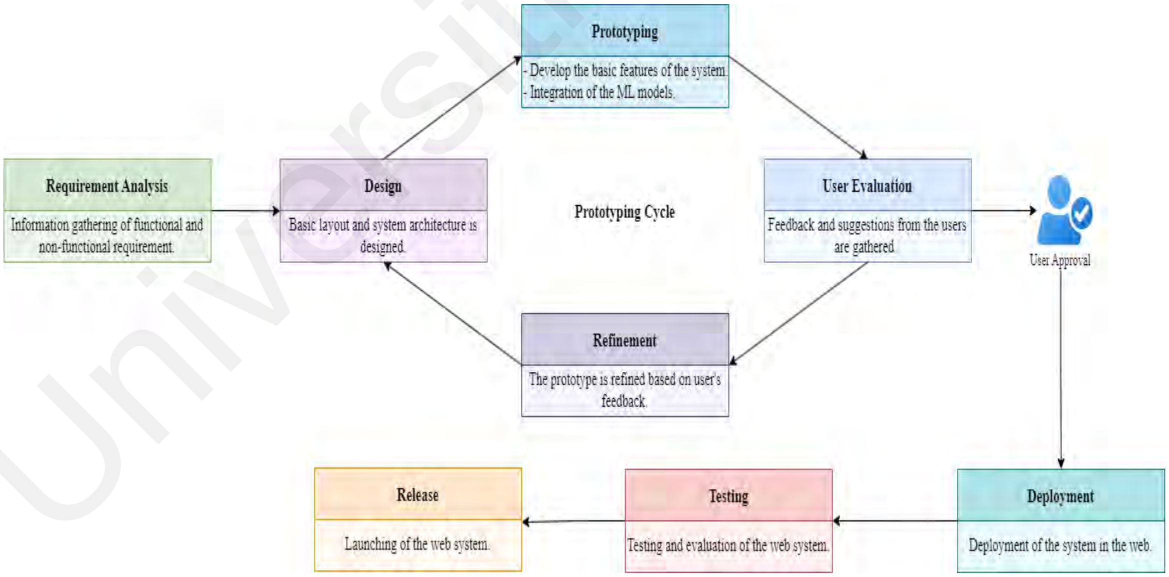

Prototyping
- Develop the basic features of the system.
- Integration of the ML models.
Requirement Analysis
Information gathering of functional and non-functional requirement.
Design
Basic layout and system architecture is designed.
Prototyping Cycle
User Evaluation
Feedback and suggestions from the users are gathered
User Approval
Refinement
The prototype is refined based on user's feedback.
Release
Launching of the web system.
Testing
Testing and evaluation of the web system.
Deployment
Deployment of the system in the web.

Figure 3.10: The prototyping cycle of the web system development.

3.15.2 Requirement Analysis

The requirement analysis is the initial stage of developing the web system for our research. This process is essential as it ensures that the web system meets the needs of the users, and it is an effective way of meeting the user needs and reducing the cost of implementation. The requirement analysis covers the area of functional requirements and non-functional requirements for this research.

Both functional and non-functional requirements describe the specific requirements of the web system must have, the difference is where the functional requirements describe about what the web system functionality instead the non-functional requirements specify on the quality attributes of the system (Altexsoft, 2022).

3.15.2.1 Functional Requirements

Functional requirements the capabilities of the system to satisfy and to be accepted by the users. The requirements are typically expressed in terms of inputs, outputs, and processes. The following table (Table 3.12) describes the functional requirements for the system prototype.

Table 3.12: Functional requirements of ACS and air pollution web system.

Functional Requirements	Descriptions
Homepage	- Serves as the main entry point to the website.
User Login page	- Only registered and verified users are allowed to use the system. - The user's detailed are maintained and stored in the database, it could only be viewed and monitored by administrator.
New User Registration page	- New user registration form, including fields for username, email, password, confirm password, and organization.
About Us page	- Contains information about the web system, such as the background of the system, purpose of the system, FAQ, and team members.

Table 3.12, continued.

Functional Requirements	Descriptions
Dashboard page	- Serves as the landing page once the user successfully login into the system, including access to all functionality of the website.
Single Site Hospital Location page	- Allows users to select a hospital and input air pollution readings in the required field. - The acquired information will then be processed by the ML model API and stored in the database. - The predicted results along with the geospatial map will be displayed to users once the API processes the information provided by the user.
Multiple Site Hospital Location page	- Users can view multiple locations and associated air pollution readings. - The system will return the predicted results on geospatial map.
ACS Hospitalization and Mortality Rate Calculator page	- The air pollutions readings as the input. - The API will process the provided input and return the predicted ACS hospitalization and ACS mortality rate. - This is a basic calculator; the data will not be stored in the database.
In-hospital Mortality Risk Calculator page	- Patient’s details, such as heart rate, ECG abnormalities and air pollution readings are required. - The data acquired will be passed to the ML API and stored in the database. - The predicted mortality risk result will be displayed once the ML is processed.

Table 3.12, continued.

Functional Requirements	Descriptions
Emergency Mortality Risk Calculator page	- Similar to 'Mortality Risk Calculator (Selected Variables)' page but requires fewer input. - This page focused on 'emergency' variables to provide the mortality risk prediction quickly.
Data Management page	- The location information can be edited, viewed, and deleted by the user. - A new location can be added and managed. - Patients' information can be viewed, updated, deleted, and downloaded from the system.

3.15.2.2 Non-functional Requirements

Non-functional requirements describe a system operation capability and constraints to improve its functionality. It specifies the quality attributes of the system to ensure the usability and effectiveness of the software system we developed. For example, the performance of the system, security, usability, and reliability. Table 3.13 below shows the non-functional requirements of the system we proposed.

Table 3.13: Non-functional requirements of ACS and Air Pollution web system.

Non-functional Requirements	Descriptions
Performance	- The system should respond quickly to users' actions. Elements such as user inputs, loading pages and processing requests that require a longer pre-loading time should be reduced and within an acceptable timeframe.
Security	- Only authorized users are allowed to access the system. No third party has the right to access the data. This ensures the privacy of the patients and prevents data breaches. - New users are required to register and login to access the web system and its functionalities. The new users are verified by the administrator.

Table 3.13, continued.

Non-functional Requirements	Descriptions
Usability	- The system is designed to be intuitive and easy to navigate. - A navigation bar is provided for the users to navigate and access other pages. - Clear instructions and guide are provided for the users.
Reliability	- The web system must be available and accessible always, with minimal downtime. - The web system should be reliable and can be accessed by various browsers including Google Chrome, Microsoft EDGE, Safari, and other web browsers. - The ML models are trained, tested, and validated before implement into the web system to ensure that the accuracy of the calculator is the same as in the developed models.
Efficiency	- The web system is expected to have sufficient processing power and storage space to ensure smooth operation. - The web system should have efficient data storage and retrieval.
Understandability	- The overall system should be easy to understand, both for users and developers. - Clear and consistent interface design, the input form is easy to understand and fill in by the users. - Long sentences in acquiring information from the users is avoided. Precise and straight-to-the-point sentences should be used. - Explanation of the site functionality is kept minimal; FAQ and upload template is provided for users as guide.

3.15.3 System Process Model

In this section, we discuss various process models used to design our web system and database. These include the workflow diagram, functional decomposition diagram, data flow diagram, data dictionary and website wireframes. Each of these models is important for a clear understanding of our system design and functionality.

3.15.3.1 Workflow Diagram

The workflow diagram gives the overall visualization of the project and system layout. It serves multiple purposes, such as tracing the system's processes, identifying, and removing unneeded or repetitive tasks, and enhancing the project's accountability and efficiency. The workflow diagram proposed in our study is depicted in Figure 3.11 below:

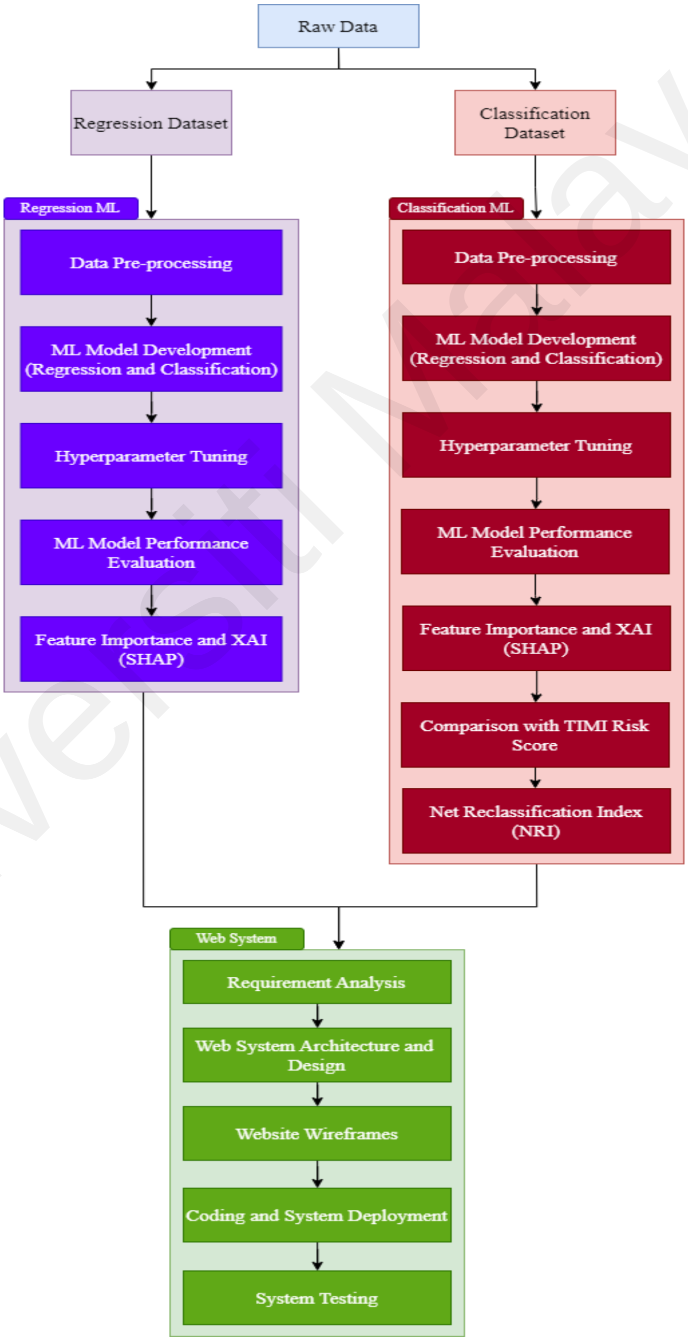

Raw Data
Regression Dataset
Classification Dataset
Regression ML
Data Pre-processing
ML Model Development (Regression and Classification)
Hyperparameter Tuning
ML Model Performance Evaluation
Feature Importance and XAI (SHAP)
Classification ML
Data Pre-processing
ML Model Development (Regression and Classification)
Hyperparameter Tuning
ML Model Performance Evaluation
Feature Importance and XAI (SHAP)
Comparison with TIMI Risk Score
Net Reclassification Index (NRI)
Web System
Requirement Analysis
Web System Architecture and Design
Website Wireframes
Coding and System Deployment
System Testing

Figure 3.11: Workflow diagram of the proposed ACS and air pollution web system.

The workflow begins with the processing of the acquired data, these processed data is applied in the ML model development, to identify the best ML model. The best performing model is selected and embedded into the web system, serving as the predictive model for users. The API development was executed using R in RStudio.

The system architecture was developed after analyzing user requirements. The user interface is then designed to incorporate the proposed system's features. The prototype is created to ensure that the system is functional and effectively aligns with the needs of the users.

The prototype is the setup and installed on the server after the users are satisfied with the prototype. The system is then presented to prospective users, and feedback is gathered via a usability testing questionnaire. Any issues that are discovered are debugged, and improvements are made based on this feedback, resulting in the completion of the final web system.

3.15.3.2 Functional Decomposition Diagram

The functional decomposition diagram (FDD) is a hierarchical method that breaks down a system into its key functions and sub-functions. Starting from the overall system function at the top, it outlines the primary functions and then further subdivides these into more detailed functionalities. This approach helps to organize the system's activities, identify overlaps, and ensure no functionality is overlooked (Inmon, et al., 2019). The FDD of our ACS and air pollution system is presented in Figure 3.12 below.

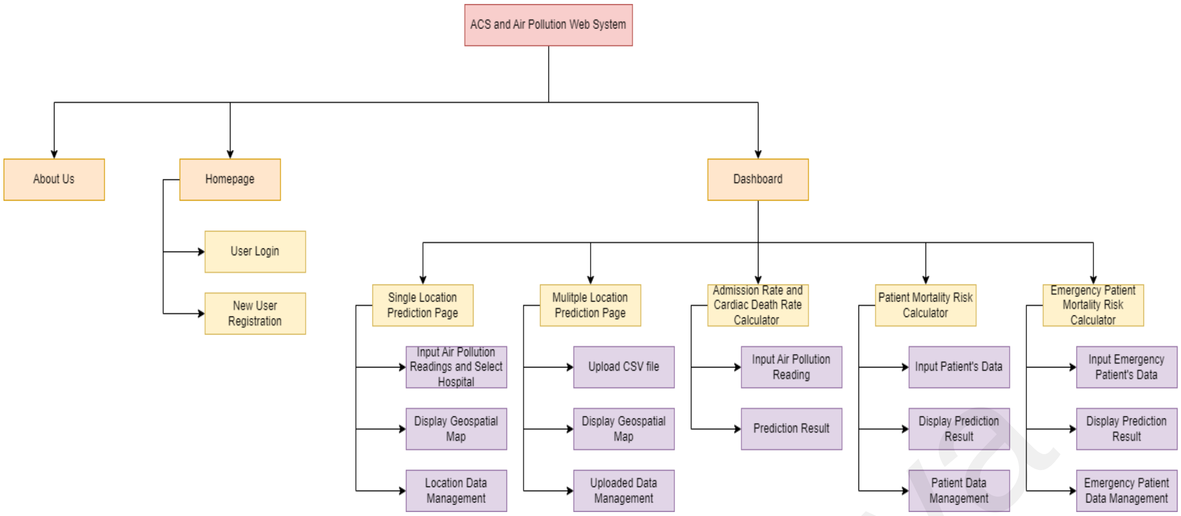

ACS and Air Pollution Web System
About Us
Homepage
Dashboard
User Login
New User Registration
Single Location Prediction Page
Mulitple Location Prediction Page
Admission Rate and Cardiac Death Rate Calculator
Patient Mortality Risk Calculator
Emergency Patient Mortality Risk Calculator
Input Air Pollution Readings and Select Hospital
Display Geospatial Map
Location Data Management
Upload CSV file
Display Geospatial Map
Uploaded Data Management
Input Air Pollution Reading
Prediction Result
Input Patient's Data
Display Prediction Result
Patient Data Management
Input Emergency Patient's Data
Display Prediction Result
Emergency Patient Data Management

Figure 3.12: Functional decomposition diagram for ACS and air pollution web system.

The top level of this system covers its overarching function: providing a platform for examining the impact of air pollution on ACS patients. This is broken down into three main second-level functions: 'About Us', 'Homepage', and 'Dashboard'. The ‘About Us’ page encompasses information regarding our study, such as the background and the purpose of this web system, the developers, and Frequently Asked Questions (FAQ).

The 'Homepage' function is further divided into two sub-functions: 'User Login' and 'New User Registration', which manage user access to the system.

The 'Dashboard' function, a crucial part of the system, is subdivided into five specific functionalities. These include pages for 'Single Location Prediction' and 'Multiple Location Prediction', and calculators for 'Admission and Mortality', 'Mortality Risk for In-Hospital ACS Patients' and 'Mortality Risk for Emergency ACS Patients'. Each of these sub-levels further encompasses minor functionalities that, collectively, contribute to the efficient operation and user experience of the web system.

By using functional decomposition, we were able to ensure a comprehensive and user-friendly design, covering all necessary components of our web system.

3.15.3.3 Data Flow Diagram

Data flow diagram (DFD) is a graphical tool that shows the flow of data in a system. The DFD includes several components which are the data flow, process, data store and entities. To depict the flow of data in the system, a context diagram and level 0 diagram are constructed.

(a) Context Diagram

A context diagram is the first level of the DFD, which contains the main process of the overall system. It is the most abstract view of a system, displaying the overview of ACS and air pollution web system. As depicted in Figure 3.13, the context diagram highlights the main process, inputs, and outputs, along with their interactions with external systems.

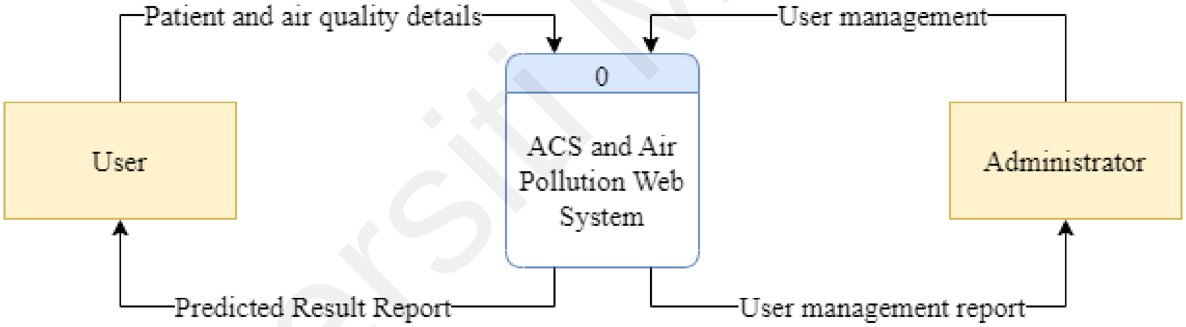


Figure 3.13: ACS and air pollution web system context diagram.

The ACS and Air Pollution web system is the core of the study; thus, it is in the center of diagram. It interacts with medical personnel who provide necessary input information, such as air pollution readings and ACS patient information. The system processes these inputs, yielding predictive results and geospatial maps as outputs. These results are then communicated back to the users completing the cycle of information exchange.

Furthermore, the web system administrator is responsible for managing the users who access the system. As a result, only authorized users can access the web system, and new registered users can only access the system with administrator approval.

(b) Level 0 Diagram

Diagram 0, shown in Figure 3.14, provides a more detailed view of the system. It expands upon the main processes, data flows, and data stores that were introduced in the context diagram. Essentially, Diagram 0 repeats and breaks down the elements of the context diagram, making it easier to understand the different parts of the system.

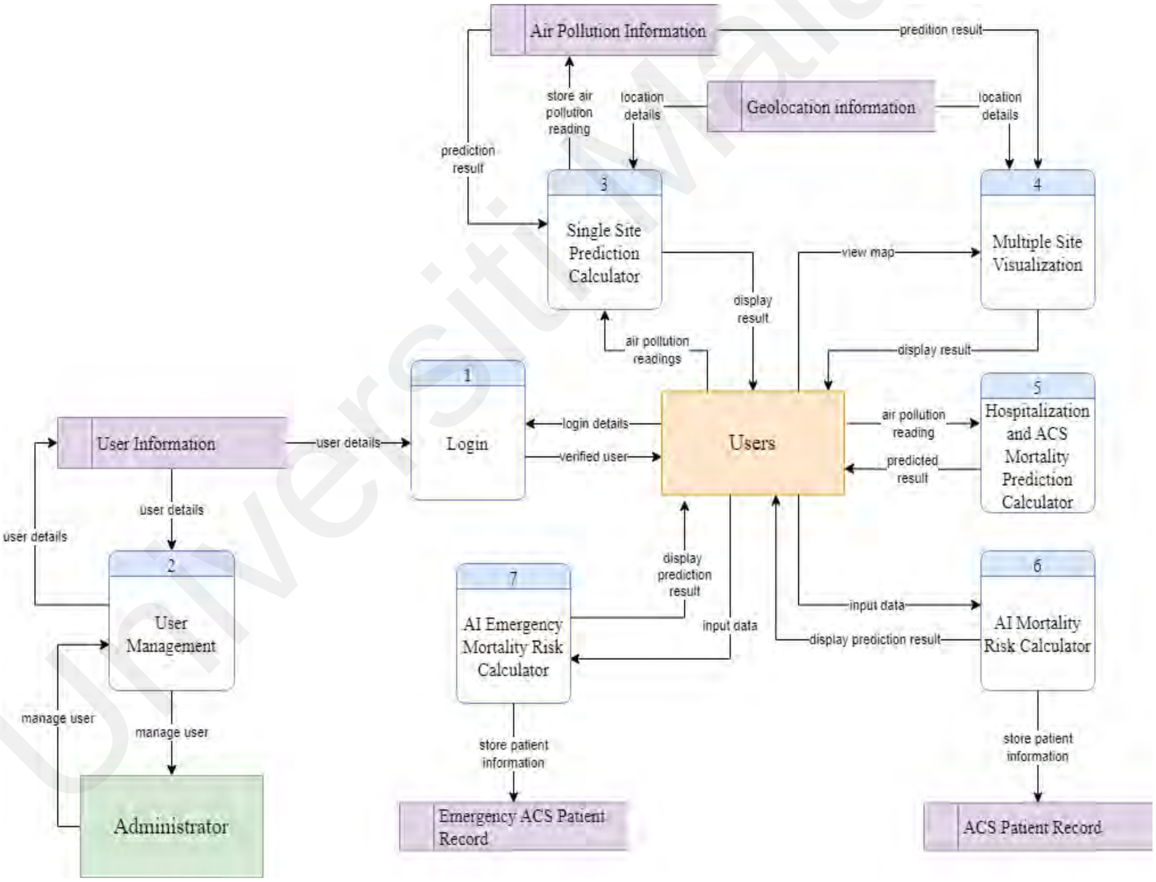


Figure 3.14: Diagram 0 for the ACS and air pollution web system.

The following walkthrough explained the DFD Diagram 0 illustrated in Figure 3.15:

1. In the initial process of DFD level 1, a registered and verified medical professional logs into the system. This process is initiated when the user inputs their login details.

Following this, the login system authenticates the provided details by cross-verifying them with the user information stored in the database. If the provided information aligns with the stored details, the login is approved, and the user gains access to the system. Once logged in, the user can access the five distinct sub-modules, each serving a unique function within the system.

2. In the single prediction web module, users input air quality readings. The system processes this data, stores it in the database, and subsequently generates prediction results. These results are then passed to another process responsible for generating a geospatial map. The system ultimately provides the user with both the prediction results and the corresponding geospatial map.
3. In the multiple site prediction web module, the system presents the user with both the prediction results and corresponding geospatial map. The data is acquired from geolocation database and air pollution database, then the system will display the multiple sites based on the input from the Single site prediction web module.
4. The ACS Hospitalization and Mortality Prediction Calculator is another key component of the system. It offers a straightforward functionality wherein the user inputs air pollution readings. The system subsequently processes these readings and promptly returns predicted results. It should be noted that this process does not involve storing information in the database.
5. Processes 6 and 7 relate with the Mortality Risk Calculator and the Emergency Risk Calculator, respectively. The process itself of these two processes is fundamentally similar, with the only difference being the amount and type of data input. The Emergency Risk Calculator takes less input data than the Mortality Risk Calculator. After users provide the required data, the system analyses it and stores it in the proper databases before creating and returning predictions to the user.

3.15.3.3 Data Dictionary

The data dictionary is a repository that describes the characteristics of the data elements stored in the database. In the context of our study, databases are used to store the user input information, the list of these databases along with their respective descriptions are as follows:

1. Hospital_location: Stores the hospital location information.
2. Hospital_admit_mortality: Stores the predicted result of the ACS hospitalization rate and mortality rate.
3. Hospital_air: Stores the process predicted readings of ACS hospitalization and mortality information.
4. Patient_sel: Stores the information for the mortality risk calculator of in-hospital patients.
5. Patient_emer: Stored the information for the mortality risk calculator of emergency patients.
6. Users: Stored the user information.

Tables 3.14 - 3.19 present the data dictionary for the web system, elaborating on the attributes and characteristics of the data elements used within these databases.

Table 3.14: Hospital_location data dictionary.

Column Name	Data Type Name	Max Length	Description
Sdp_id (PK)	int	11	Source Data Provider ID
Hospital_name	varchar	50	Name of hospital
Hospital_state	varchar	40	State of the hospital located
Lat	varchar	30	Latitude
lng	varchar	30	Longitude

Table 3.15: Hospital_admit_mortality data dictionary.

Column Name	Data Type Name	Max Length	Description
Sdp_id	int	11	Source Data Provider ID
Hospital_name	varchar	50	Name of hospital
Hosp_date	date		Recorded input date
admit	int	11	Predicted ACS hospitalization rate
mortality	Int	11	Predicted ACS mortality rate

Table 3.16: Hospital_Air data dictionary.

Column Name	Data Type Name	Max Length	Description
Id (PK)	Int	11	Auto Increment ID
Sdp_id	int	11	Source Data Provider ID
Hospital_name	varchar	50	Hospital Name
date	date		Recorded date
Nox	Float		Nitrogen Oxides Reading
SO_2	Float		Sulphur Dioxide reading
O_3	Float		Ozone reading
PM10	Float		Particulate Matter 10 reading

Table 3.17: In-hospital patient data dictionary

Column Name	Data Type Name	Max Length	Description
Id (PK)	Int	11	Auto Increment ID
Username	int	11	Username of the login user
Date	Date		Record date of input
Pic	Int	25	Patient Identification ID
ptageatnotification	Int	11	Patient Age
Canginapast2wk	Int	11	Chronic angina past 2 weeks
Heartrate	float		Heart Rate
Killipclass	Int	11	Killip class
Hdlc	float		High density lipoprotein cholesterol

Table 3.17, continued.

Column Name	Data Type Name	Max Length	Description
Ldlc	float		Low density lipoprotein cholesterol
Ldlc	float		Low density lipoprotein cholesterol
fbg	float		Fast Blood Glucose
Ecgabnormtypestelev1	Int	11	ECG abnormal ST-elevation Type 1
Cardiaccath	Int	11	cardiac catheterization
cabg	Int	11	Coronary artery bypass graft
Statin	Int	11	Statin medication
Lipidla	Int	11	Lipid Lower Agent medication
Oralhypogly	Int	11	Oral Hypoglycemic medication
Antiarr	Int	11	Antiarrhythmics medication
Nox	Float		Nitrogen Oxides Reading
SO_2	Float		Sulphur Dioxide reading
O_3	Float		Ozone reading
PM10	Float		Particulate Matter 10 reading
mortality	Float		The probability of the ACS patient mortality.
Mortality_percentage	int	25	The percentage of patient mortality risk.
Real_inhosp	varchar	25	Update by user whether the patient is 'Alive' or 'Dead'
Remarks	varchar	1000	
Last_updated	Date		Current time stamp

Table 3.18: Emer_patient data dictionary.

Column Name	Data Type Name	Max Length	Description
Id (PK)	Int	11	Auto Increment ID
Username	int	11	Username of the login user
Date	Date		Record date of input
Pic	Int	25	Patient Identification ID
ptageatnotification	Int	11	Patient Age
Canginapast2wk	Int	11	Chronic angine past 2 weeks
Heartrate	float		Heart Rate
Killipclass	Int	11	Killip class
Ecgabnormtypestelev1	Int	11	ECG abnormal ST-elevation Type 1
Statin	Int	11	Statin medication
Lipidla	Int	11	Lipid Lower Agent medication
Oralhypogly	Int	11	Oral Hypoglycemic medication
Antiarr	Int	11	Antiarrhythmics medication
NOx	Float		Nitrogen Oxides Reading
SO_2	Float		Sulphur Dioxide reading
O_3	Float		Ozone reading
PM10	Float		Particulate Matter 10 reading
Prob_emer	Float		The probability of the emergency ACS patient mortality.
Mortality_percentage	int	25	The percentage of emergency patient mortality risk.
Real_inhosp	varchar	25	Update by user whether the patient is 'Alive' or 'Dead'
Remarks	varchar	1000	
Last_updated	Date		Current time stamp

Table 3.19: Users data dictionary.

Column Name	Data Type Name	Max Length	Description
Id (PK)	Int	11	Auto generated ID
Date	Date		Date of registered
Email	varchar	255	Email of the user
Username	varchar	255	Username of the user
Password	varchar	255	Password used
Registered	varchar	5	Approval for the user to access the system

3.15.4 User Interface Design and Human Computer Interaction

The design of the user interface (UI) is a fundamental aspect of web system development, where it affects how the user interacts with the system and their experience. A well-designed UI contributes significantly to the overall usability of the system, enhancing user satisfaction and engagement.

The eight golden rules proposed by (Shneiderman & Plaisant, 2004) serves as our guideline and starting point in design the user interface for the ACS and air pollution system. The 8 golden rules are as listed in Table 3.20 below:

Table 3.20: 8 golden rules for designing ACS and air pollution user interface.

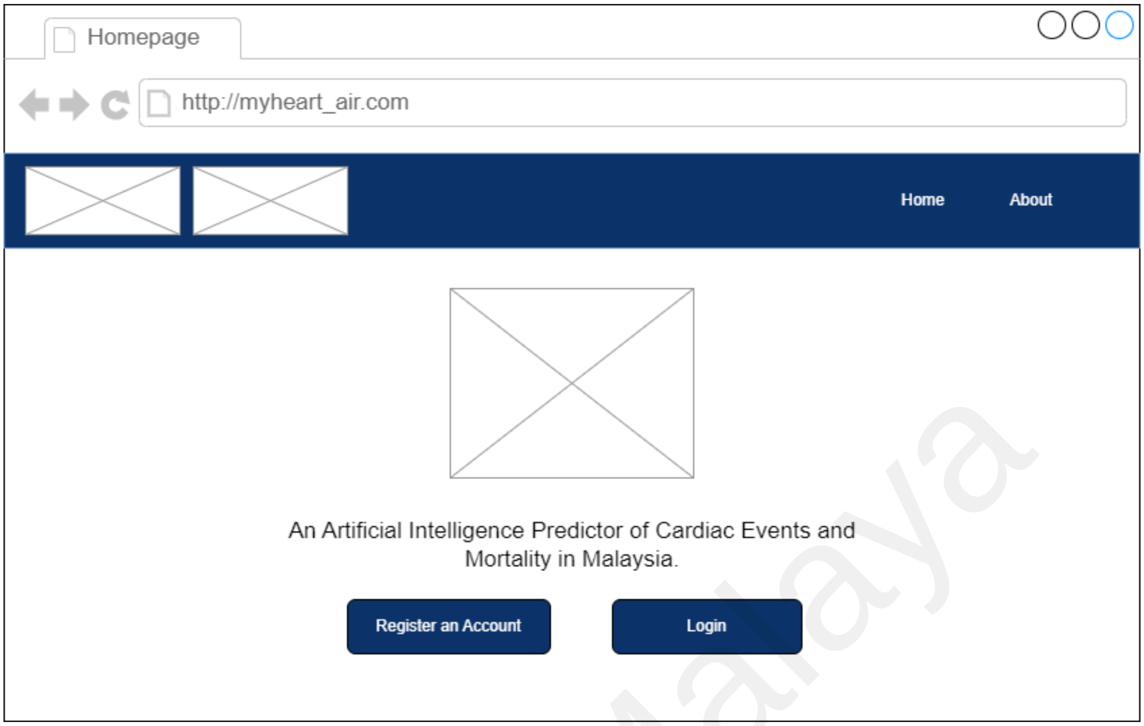
8 Golden Rules	Descriptions
Strive for consistency.	The design across the system is kept uniform, we used the same color scheme, typography, and button styles throughout the web system.
Cater to universal usability.	Ensure the system user friendly, where the design of our system is simple and intuitive.
Offer informative feedback.	Clear responses to user actions, when a user interacts with the system, the system will receive immediate and clear messages, such as "Data successfully uploaded".
Design dialogs to yield closure.	Group actions into task units with clear beginnings and ends. Each user task, such as data input, has a distinct start and endpoint, guiding the users go through the entire process.

Table 3.20, continued.

8 Golden Rules	Descriptions
Prevent errors.	Minimize user errors by including warnings, and set requirements of the data input, and the range of data input, reducing the risk of unintentional mistakes.
Permit easy reversal of actions.	Users can easily modify their input data without any adverse effects to the system's function.
Support internal locus of control.	Users have complete control over all actions while navigating through the system.
Reduce short-term memory load.	Relieve user's memory load by providing prompts on each data input in each field, make use of visual aids.

The user interface design of the ACS and air pollution system followed the eight golden rules of interface design, which enhanced usability and overall system quality. Though the 8 golden rules are introduced in 2004, it is widely used in improving the usability and quality of UI design (Aottiwerch & Kokaew, 2017; Masmuzidin & Aziz, 2019).

The wireframes of our web system, shown in Figures 3.15 to 3.27. However, the design is subject to change based on the user feedback and evolving needs.



An Artificial Intelligence Predictor of Cardiac Events and Mortality in Malaysia.

Figure 3.15: Web system homepage website wireframe design.

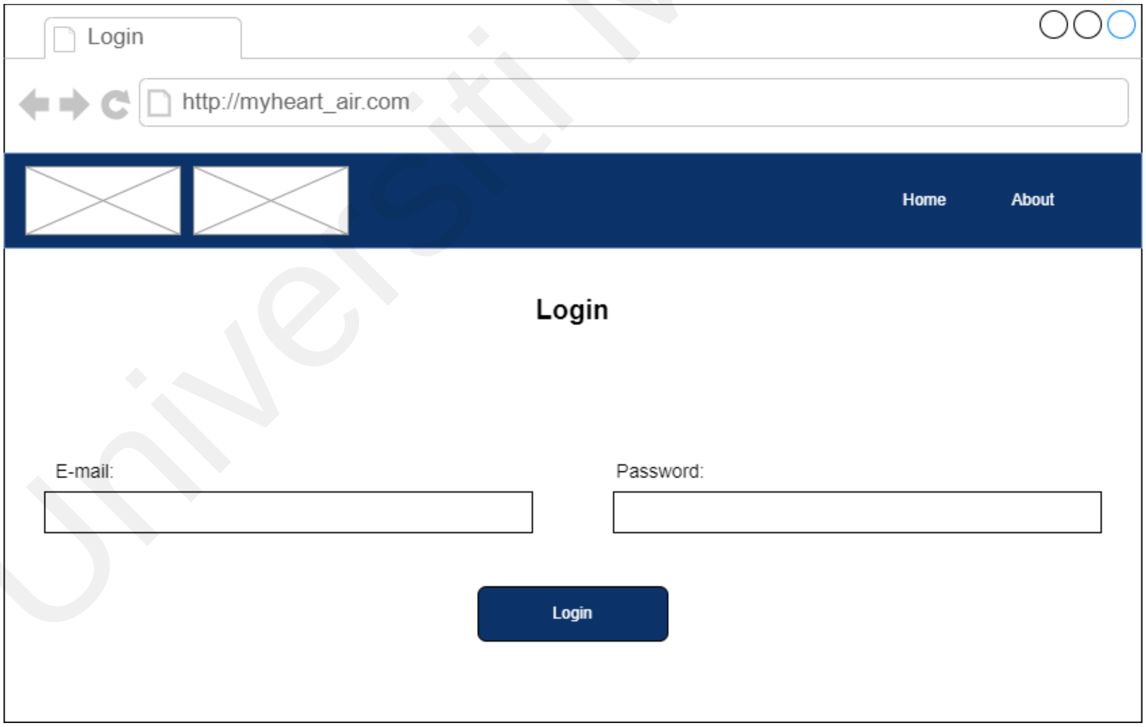


Figure 3.16: Login page website wireframe design.

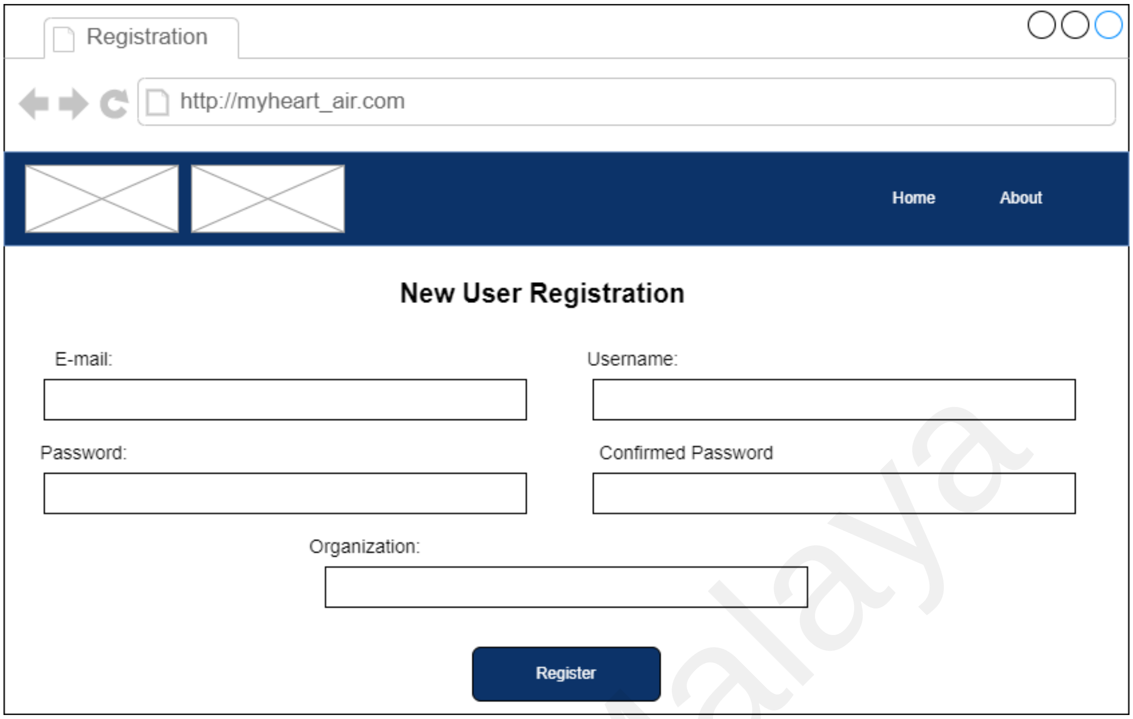
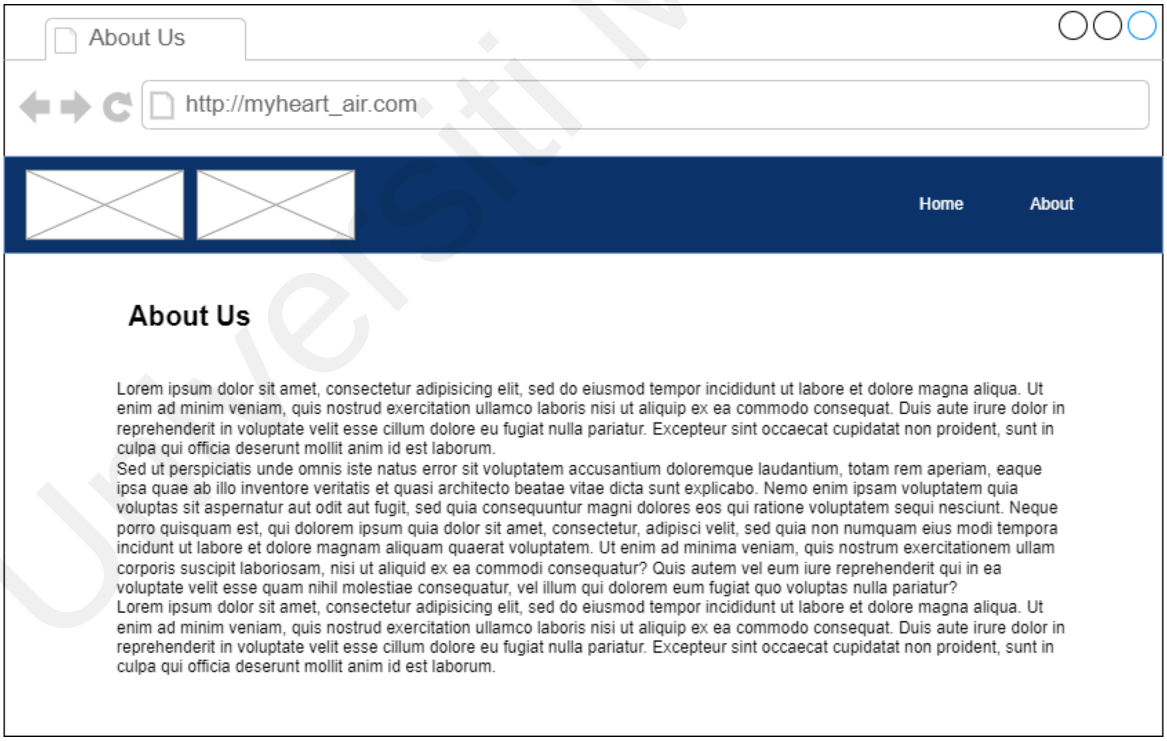


Figure 3.17: New user registration website wireframe design.



Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.
Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium doloremque laudantium, totam rem aperiam, eaque ipsa quae ab illo inventore veritatis et quasi architecto beatae vitae dicta sunt explicabo. Nemo enim ipsam voluptatem quia voluptas sit aspernatur aut odit aut fugit, sed quia consequuntur magni dolores eos qui ratione voluptatem sequi nesciunt. Neque porro quisquam est, qui dolorem ipsum quia dolor sit amet, consectetur, adipisci velit, sed quia non numquam eius modi tempora incidunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim ad minima veniam, quis nostrum exercitationem ullam corporis suscipit laboriosam, nisi ut aliquid ex ea commodi consequatur? Quis autem vel eum iure reprehenderit qui in ea voluptate velit esse quam nihil molestiae consequatur, vel illum qui dolorem eum fugiat quo voluptas nulla pariatur?
Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Figure 3.18: About Us website wireframe design.

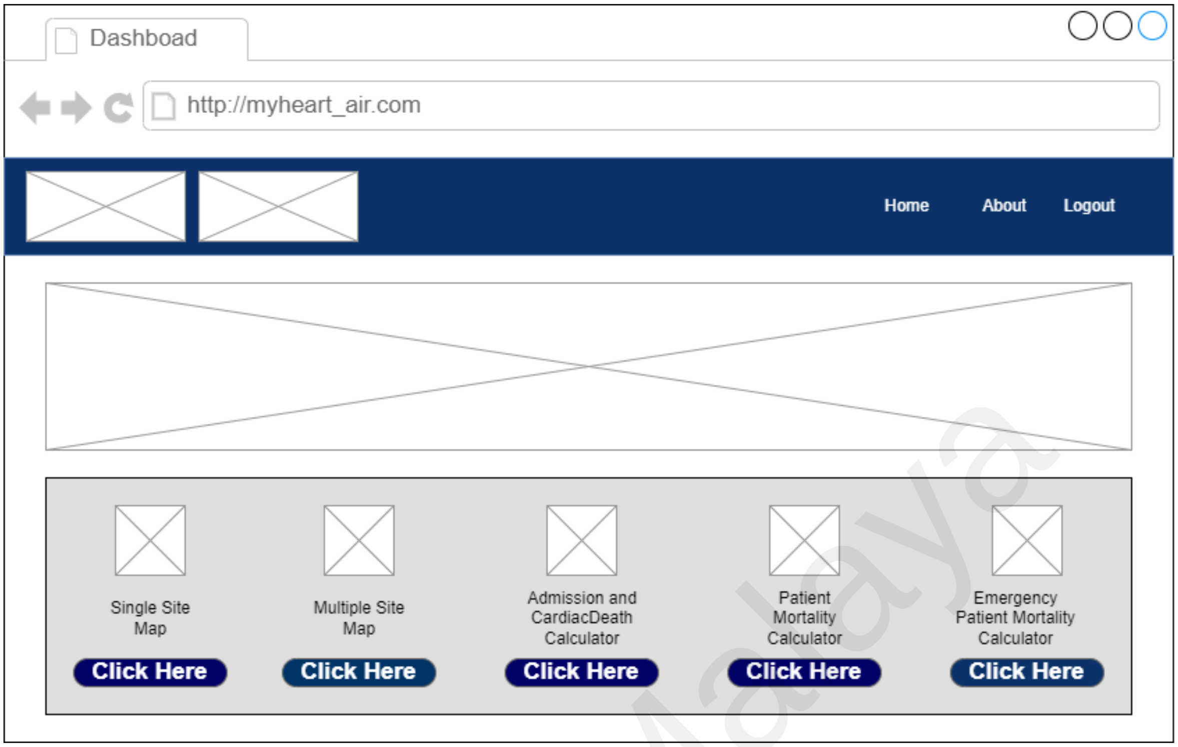
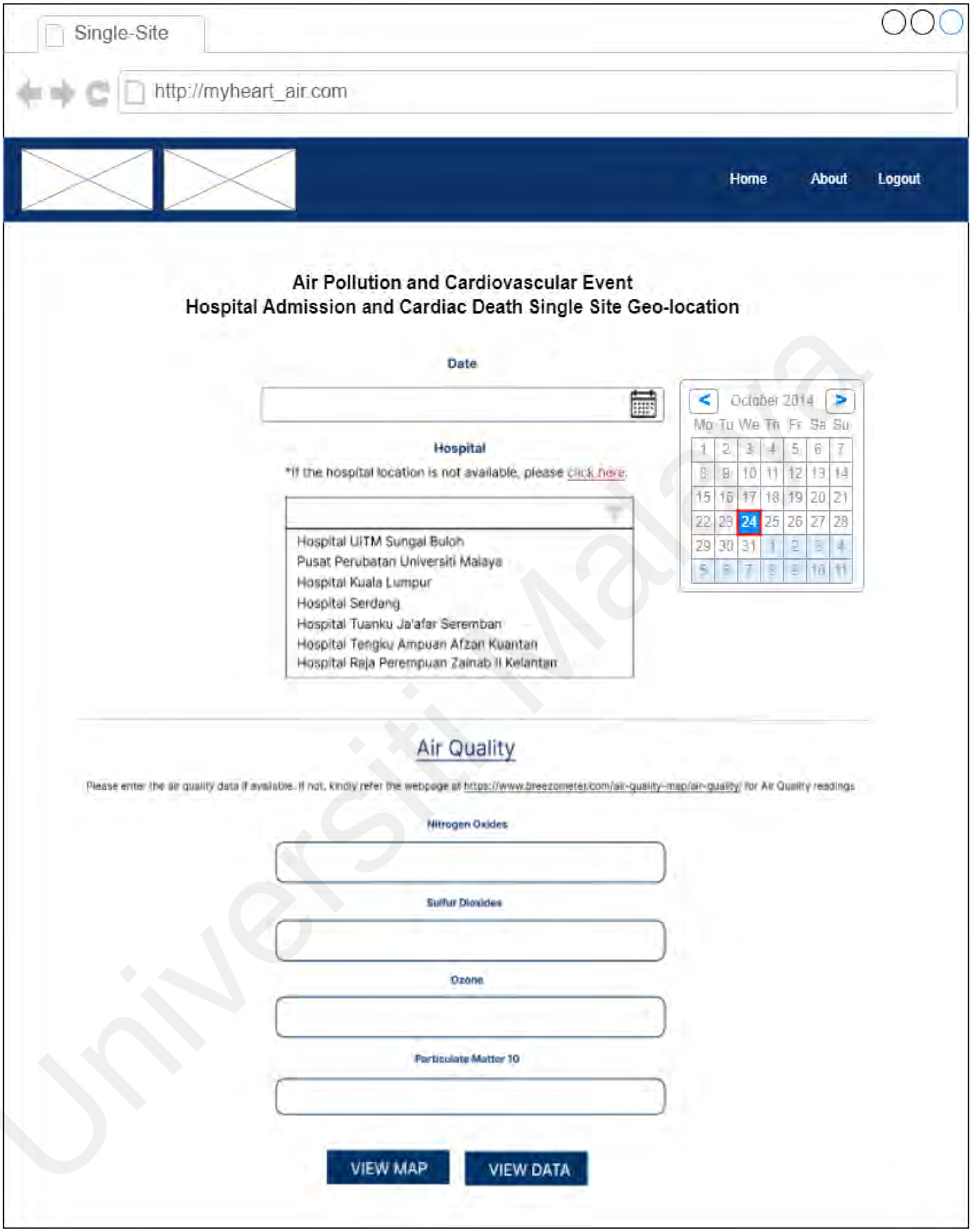


Figure 3.19: Dashboard website wireframe design. once the user successfully registered and login.



Air Pollution and Cardiovascular Event
Hospital Admission and Cardiac Death Single Site Geo-location

*If the hospital location is not available, please click here.

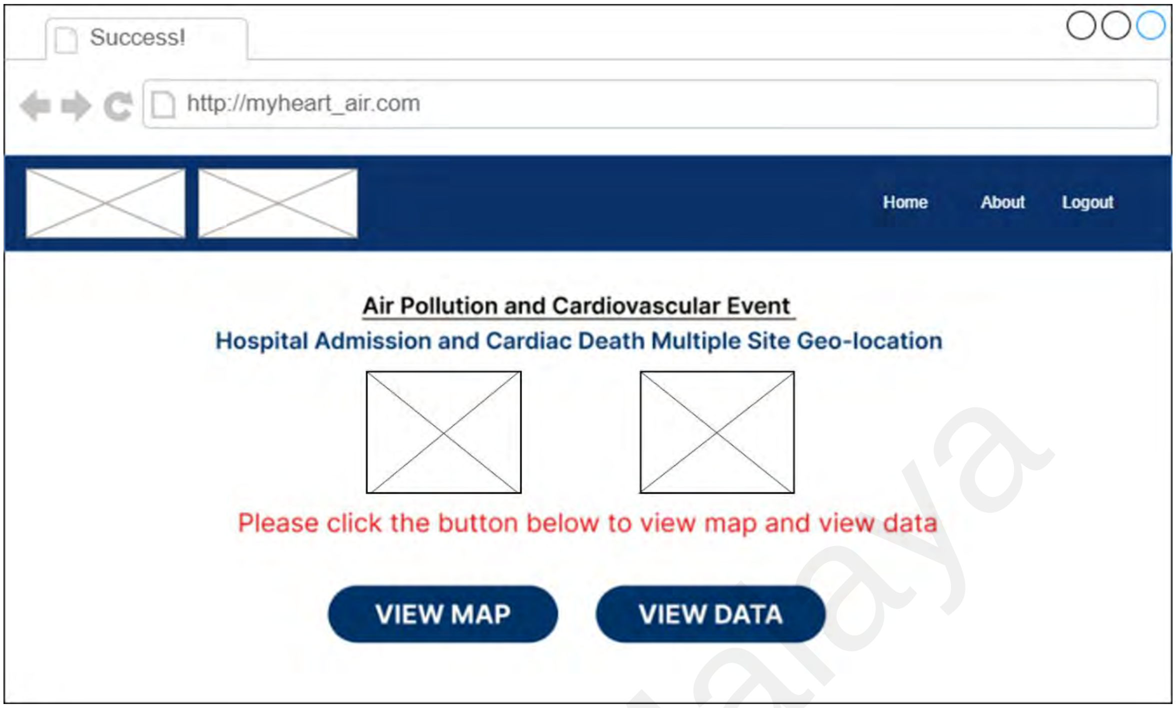
- Hospital UITM Sungai Buloh
- Pusat Perubatan Universiti Malaya
- Hospital Kuala Lumpur
- Hospital Serdang
- Hospital Tuanku Ja'afar Seremban
- Hospital Tengku Ampuan Afzan Kuantan
- Hospital Raja Perempuan Zainab II Kelantan

October 2014

Mo	Tu	We	Th	Fr	Sa	Su
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	1	2	3	4
5	6	7	8	9	10	11

Please enter the air quality data if available. If not, kindly refer the webpage at https://www.breezometer.com/air-quality-map/air-quality/ for Air Quality readings

Figure 3.20: Single site data input website wireframe design.



Air Pollution and Cardiovascular Event

Hospital Admission and Cardiac Death Multiple Site Geo-location

Please click the button below to view map and view data

Figure 3.21: Multiple site view map and view data website wireframe design.

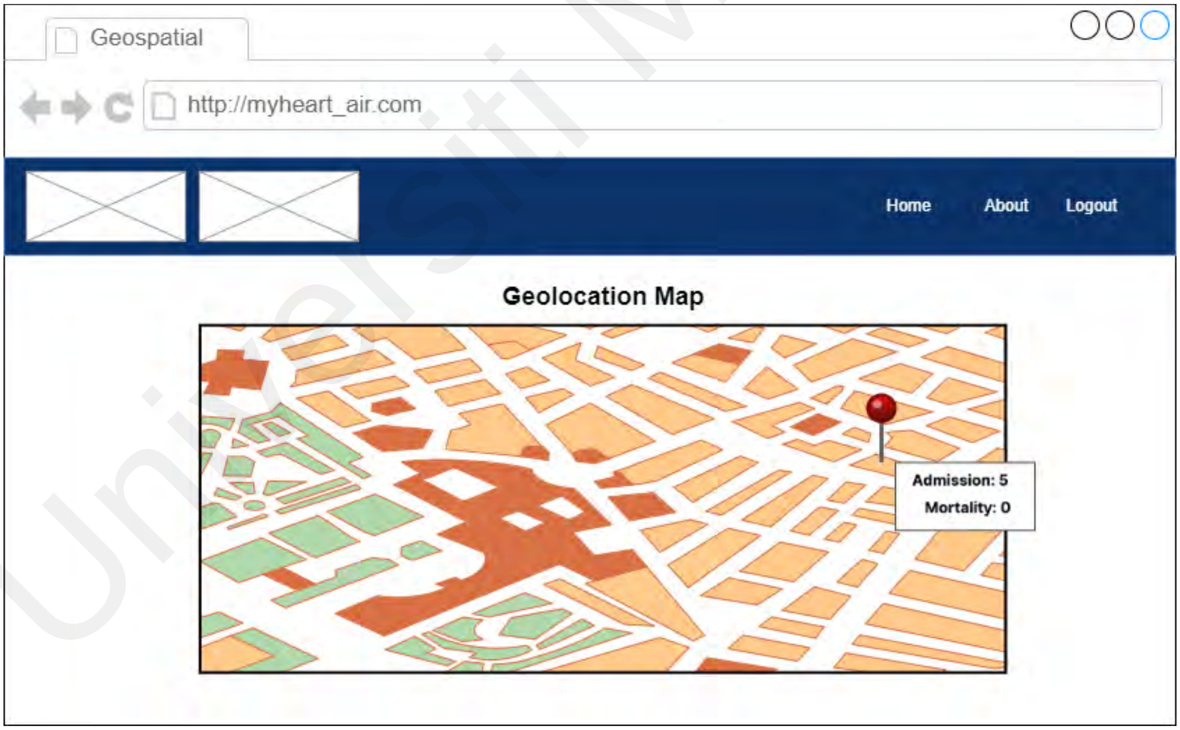
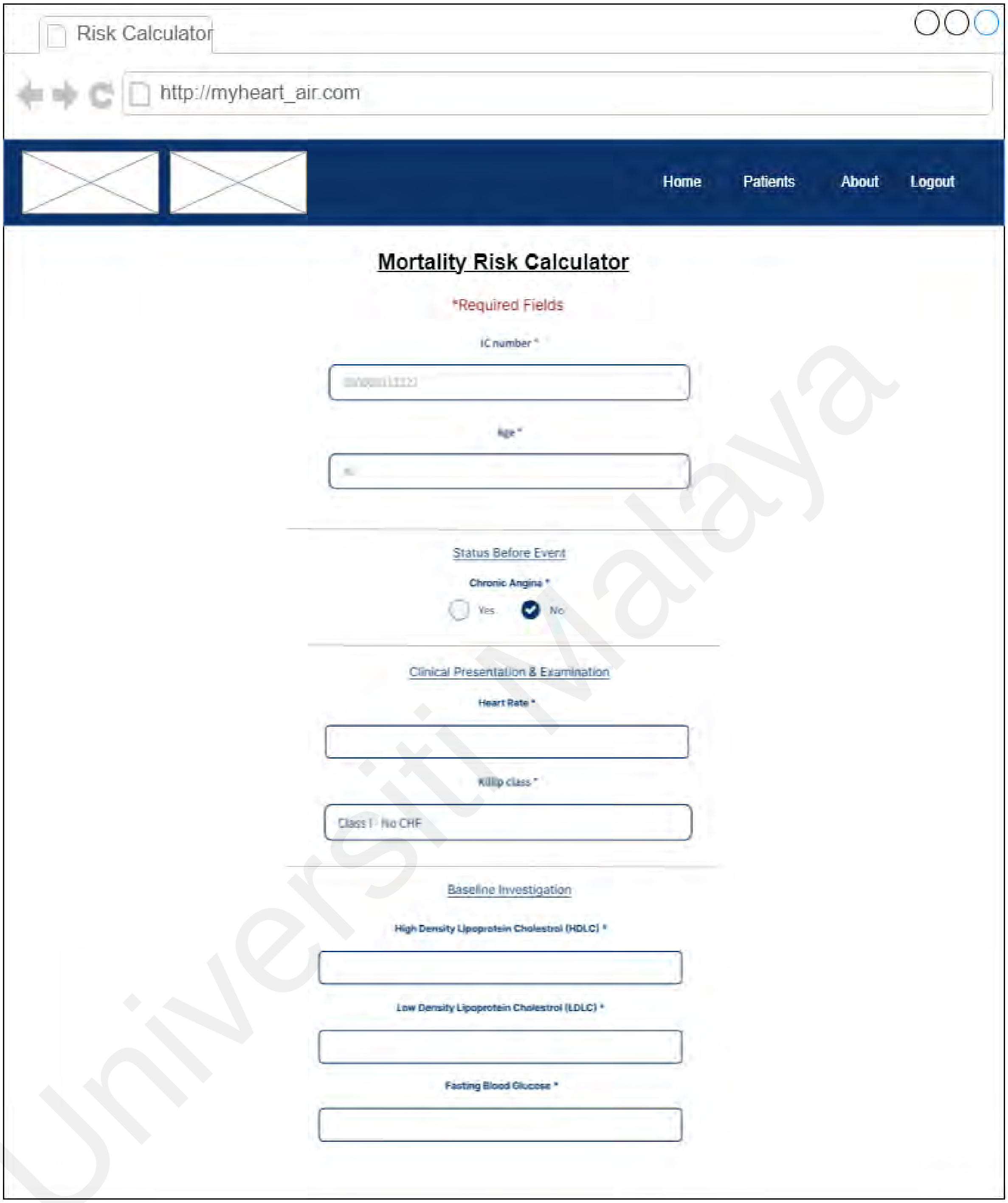


Figure 3.22: Geospatial map visualization page website wireframe design.



Mortality Risk Calculator

*Required Fields

IC number *

Age *

Status Before Event

Chronic Angina *

○ Yes ● No

Clinical Presentation & Examination

Heart Rate *

Killip class *

Class I - No CHF

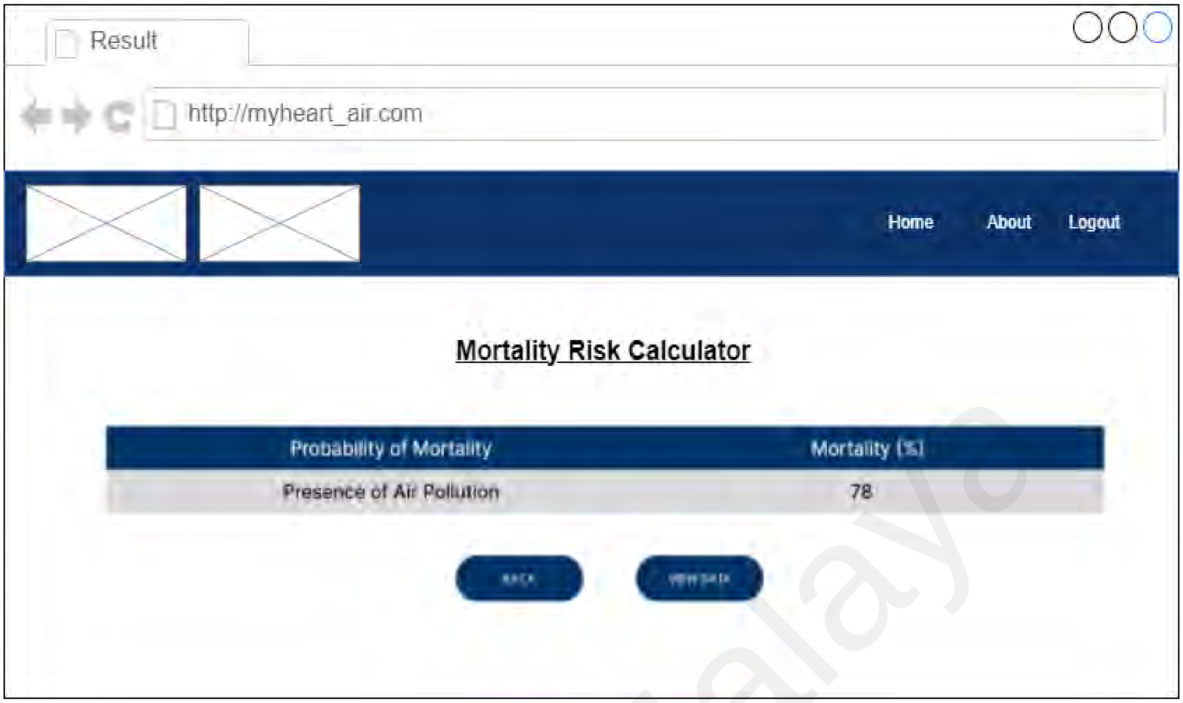
Baseline Investigation

High Density Lipoprotein Cholestrol (HDLC) *

Low Density Lipoprotein Cholestrol (LDLC) *

Fasting Blood Glucose *

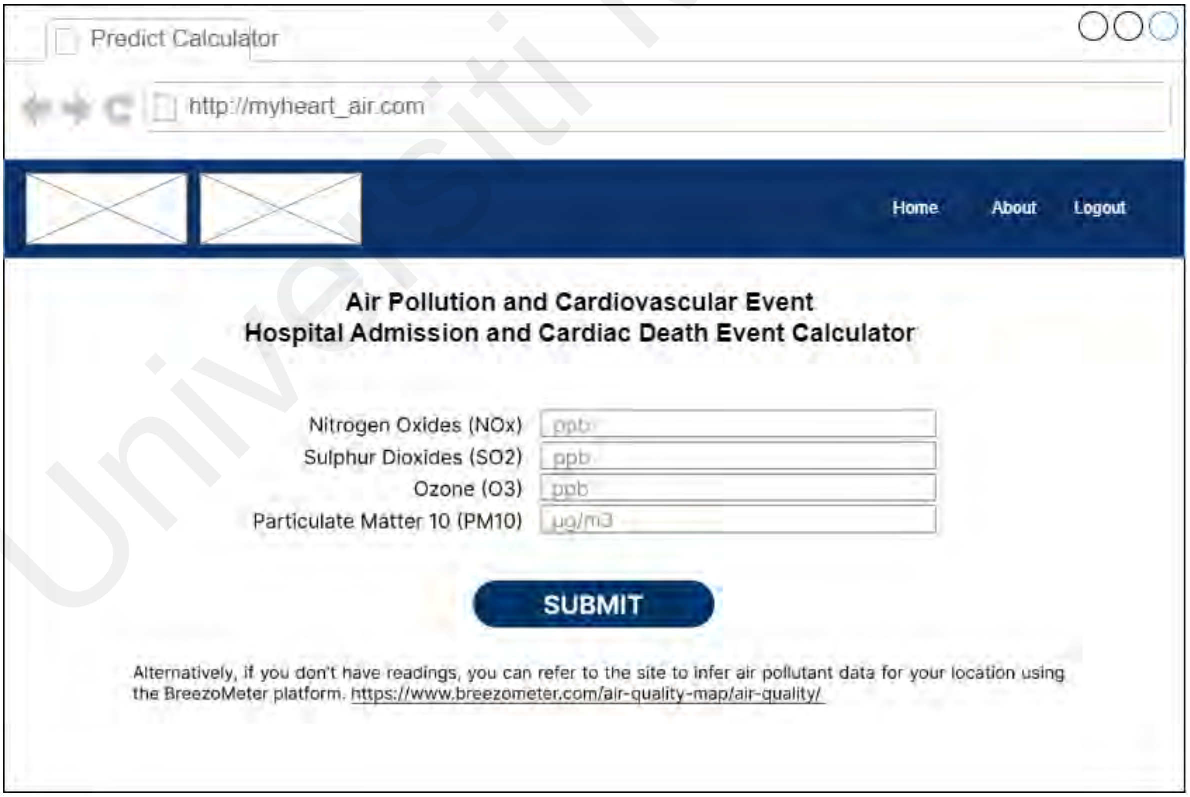
Figure 3.23: Mortality risk calculator input page website wireframe design.



Mortality Risk Calculator

Probability of Mortality	Mortality (%)
Presence of Air Pollution	78

Figure 3.24: Mortality risk result page website wireframe design. after the user provided the required input.

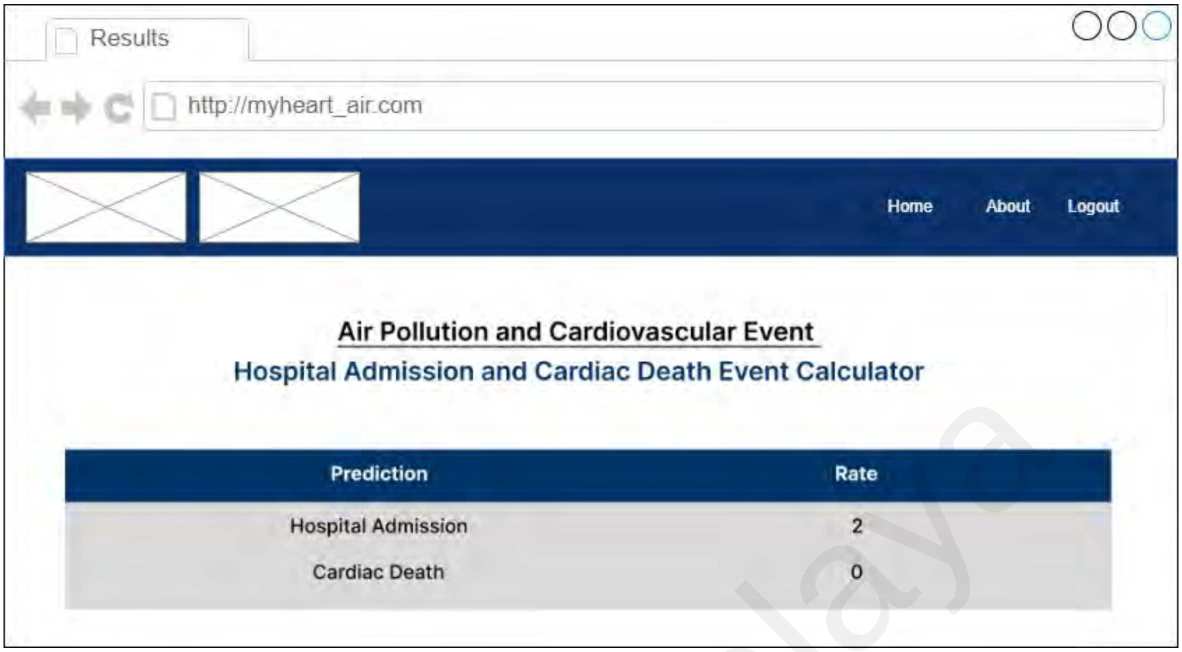


Air Pollution and Cardiovascular Event
Hospital Admission and Cardiac Death Event Calculator

Nitrogen Oxides (NOx) ppb
Sulphur Dioxides (SO2) ppb
Ozone (O3) ppb
Particulate Matter 10 (PM10) µg/m3

Alternatively, if you don't have readings, you can refer to the site to infer air pollutant data for your location using the BreezoMeter platform. https://www.breezometer.com/air-quality-map/air-quality/

Figure 3.25: ACS hospitalization and ACS mortality event calculator website wireframe design.

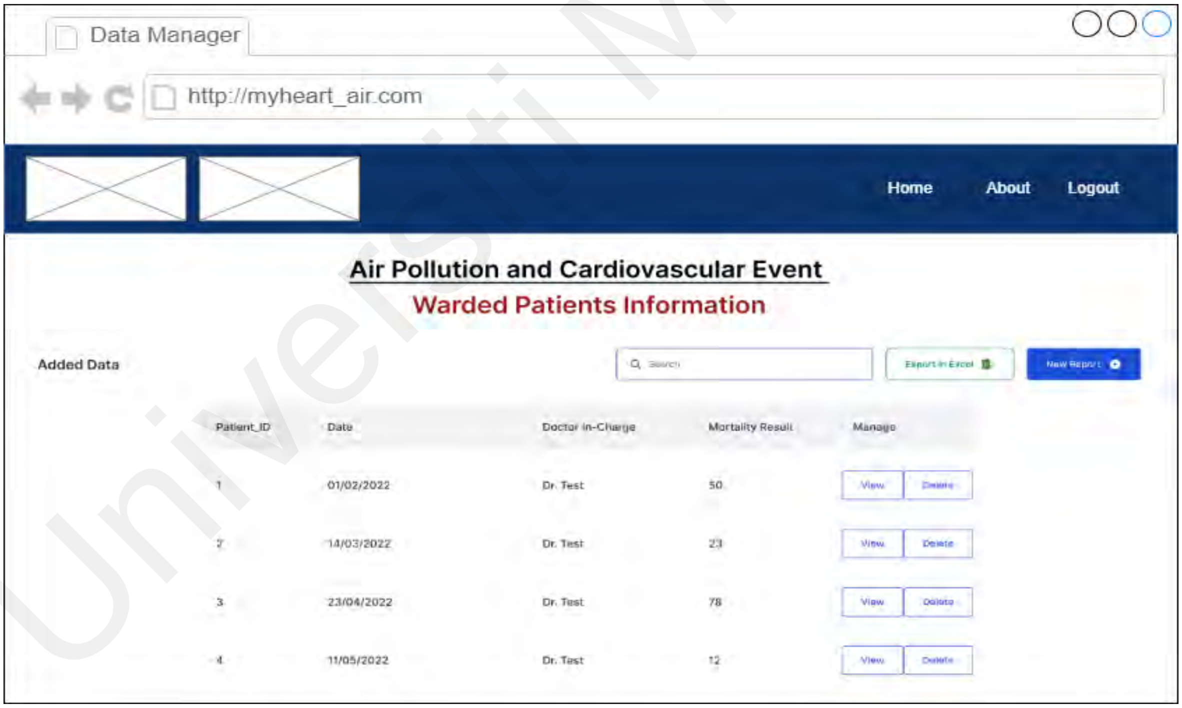


Air Pollution and Cardiovascular Event

Hospital Admission and Cardiac Death Event Calculator

Prediction	Rate
Hospital Admission	2
Cardiac Death	0

Figure 3.26: ACS hospitalization and ACS mortality event calculator display result website wireframe design.



Air Pollution and Cardiovascular Event

Warded Patients Information

Added Data

Patient_ID	Date	Doctor in-Charge	Mortality Result	Manage
1	01/02/2022	Dr. Test	50	View Delete
2	14/03/2022	Dr. Test	23	View Delete
3	23/04/2022	Dr. Test	78	View Delete
4	11/05/2022	Dr. Test	12	View Delete

Figure 3.27: Data management website wireframe design that allows user to manage data.

3.15.5. Development Environment

This section outlines the hardware software requirement that was used to develop the ACS and air pollution web system. Table 3.21 below shows the hardware and software specifications in this study.

Table 3.21: Hardware and software requirements

Hardware Requirement	Descriptions
Edition	Windows 11 Professional
System Type	64-bit operating system, x64-based processor
Processor	AMD Ryzen 7 5700U with Radeon Graphics (1.80 GHz)
Installed RAM	8.00 GB
Software Requirement	**Descriptions**
Data Processing	Microsoft® Excel® for Microsoft 365 MSO (Version 2304 Build 16.0.16327.20200)
Diagrams and flowcharts	Draw.io Figma
Statistical Analysis	IBM SPSS Statistics 26
Machine Learning Model Development	Language: R Coding Environment: RStudio 2023.03.1 +446
Machine Learning Model Analysis	Language: Python Coding Environment: Jupyter Notebook
Coding environment	Notepad++ 7
Coding environment	VS Code 1.78
PHP Development Environment	XAMPP 8.2.4

3.15.5 Machine Learning Implementation

The key component of this study is the integration of ML into the web system. After determining the ML model that delivered the best performance, the model was saved and serialized into RDS format using the saveRDS function. This format enables efficient storage and retrieval of the ML model.

Subsequently, the 'plumber' package in R was utilized to integrate the ML models into the web-based environment. The plumber package allows developers to create web APIs directly from R scripts, serving as a pivotal tool in integrating and exposing static ML models as dynamic, web-accessible resources. Once the ML models were loaded, the plumber package transformed functions into accessible APIs, creating routes that corresponded to specific R functions. In this study, four models were loaded to perform prediction functions, forecasting the ACS hospitalization rate, ACS mortality rate, and ACS mortality risk for both in-hospital and emergency patients.

Finally, to facilitate interactive testing and documentation of the API, Swagger (OpenAPI) was employed. This tool provided an essential step towards ensuring the system's functionality and the successful integration of the ML model. Once the API was prepared and operational, it was made accessible for user interaction and testing via Swagger's user-friendly interface. This pivotal stage allowed us to confirm that the ML model was correctly integrated and delivering the expected outputs.

3.15.6 System Testing

The testing phase is the last phase before launching the web system. It is an important stage for system validation, and to ensure the deliverables adhere to the design specification. It facilitates the identification of defects that may surface upon complete system assembly and integration. Besides, testing is conducted to ensure all the modules are functioning well and integrate with other components. In the context of the ACS and air pollution system examined in this study, the testing phase encompassed unit testing, system testing, integration testing and acceptance testing listed down below:

1. Unit Testing

   Each module is subjected to unit testing to ensure its functionality is accurate and bug-free. Each element of the module is tested to ensure that the source code for the module is functional. When an error arises, it must be addressed.

In this study, unit testing is run through the login functionality, data input module, prediction calculation module, geospatial visualization module, data management module and result display module to ensure each module is well-functioned.

2. System Testing

System testing is carried out after unit testing is completed. System testing aims to ensure that all the system modules can seamlessly interoperate, thereby functioning. During testing, an error is detected, the affected module is debugged and tested again. The system is fed with input data to determine whether the information processing corresponds to the correct output, thereby verifying that the system performance adheres to the specified parameters.

3. Integration Testing

Upon completion of system testing, the subsequent phase is integration testing. The aim of this stage is to ensure that the individual modules of the ACS and Air Pollution system can work with the existing system error-free. In the context of this integration testing, specific modules, including the geospatial mapping for ACS hospitalization rates and ACS mortality rates, along with the various risk calculators, are integrated into the entire web system. A key aspect of this testing phase is to validate that these modules can accurately retrieve and store data within the system's database.

4. Acceptance Testing

Acceptance testing represents the final stage of system testing, functioning as a quality assurance process to verify that the developed system aligns with end-user expectations, in terms of both functional and non-functional requirements. This involves presenting the complete web system to users, providing an overview of its functionality, and requesting them to explore and evaluate the system. User feedback and suggestions collected via a system evaluation form are then analyzed, serving as valuable insights for rectifying issues and enhancing existing features. In the case of our web system, the acceptance test employs the System Usability Scale (SUS), a metric established by John Brooke in 1986, the details of which are elaborated in section 3.15.7.1.

3.15.7.1 System Usability Scale (SUS)

The System Usability Scale (SUS) developed by John Brooke in 1986 is used as the acceptance test for our ACS and Air Pollution system. It is a low-cost assessment, fast and reliable to measure the usability in the system which only comprises 10 questions (Brooke, 1986). It is the most widely used standardized questionnaire for the assessment of perceived usability (Lewis, 2018). Table 3.22 presents the comparison of original SUS questionnaire from (Brooke, 1986) and the modified SUS statements that suits our study.

Table 3.22: The original SUS statements by Brooke (1986) and edited SUS statements.

Original SUS Statements	Edited SUS Statements
I think that I would like to use this system frequently.	I think that I would like to use ACS and Air Pollution system frequently.
I found the system unnecessarily complex.	I found that ACS and Air Pollution system unnecessarily complex.
I thought the system was easy to use.	I thought the ACS and Air Pollution system was easy to use.
I think that I would need the support of a technical person to be able to use this system.	I think that I would need the support of a technical person to be able to use this ACS and Air Pollution system.
I found the various functions in this system were well integrated.	I found the various functions in this ACS and Air Pollution system were well integrated.
I thought there was too much inconsistency in this system.	I thought there was too much inconsistency in this ACS and Air Pollution system.
I would imagine that most people would learn to use this system very quickly.	I would imagine that most people would learn to use this ACS and Air Pollution system very quickly.
I found the system very cumbersome to use.	I found the ACS and Air Pollution system very cumbersome (awkward) to use.

Table 3.22, continued.

Original SUS Statements	Edited SUS Statements
I felt very confident using the system.	I felt very confident using the ACS and Air Pollution system.
I needed to learn a lot of things before I could get going with this system.	I needed to learn a lot of things before I could get going with this ACS and Air Pollution system.

The SUS consists only of 10 questions which are scored on a 5-point scale of the strength of agreement. The range goes from “strongly agree’ to ‘strongly disagree” and because the statements fluctuate between positive and negative, additional attention must be used when responding to the survey.

The users will rank each of the questions as the following, with the score of 1 indicating “Strongly Disagree”, 2 indicates “Disagree”, 3 indicates “Neutral”, follow by a score of 4 indicating “Agree” and 5 indicating “Strong Agree”. The scores are then converted into numbers and calculated the usability score using SUS. According to Bangor, et al. (2009), the SUS score acceptability ranges from 70 and above, according to Figure 3.28 of the SUS score shown below:

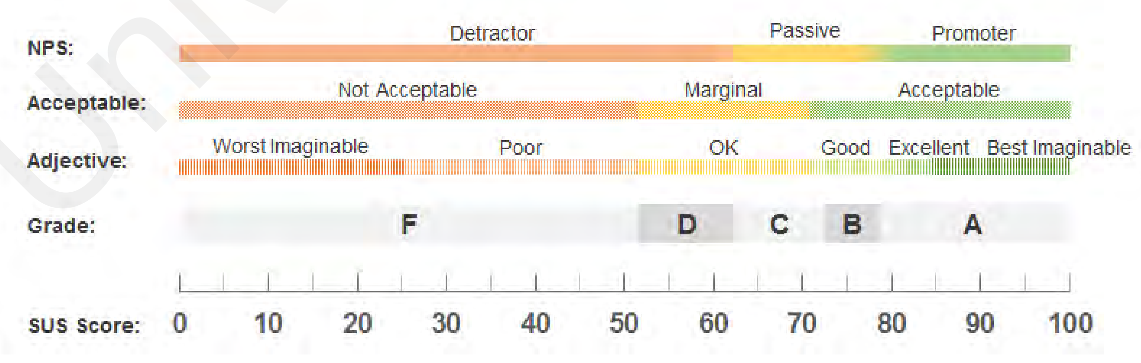


Figure 3.28: SUS scores grade rankings (Photo sourced from Bangor, et al., 2009)

SUS is chosen as a usability test based on its wide advocacy, its relatively quick processing time, where the respondents can give rapid feedback and comments, as an outcome of which the information collected is processed quickly. SUS is versatile and its wide application for various programs and application systems. The SUS score can be interpreted easily, and improvements can be made to improve the system's performance (Bhat, 2018). In Bangor, et al. (2009) study, it was found that SUS is highly reliable (alpha=0.91) and useful in wide range of tasks based on the results of 2324 SUS surveys collected from 2016 usability experiments over a decade.

The users are encouraged to explore and navigate through the system before completing the questionnaire to provide an accurate usability evaluation. The SUS questionnaire was created using Google Forms. The results of the SUS questionnaire are then analyzed and discussed in Chapter 4 "Results" and Chapter 5 “Discussion”. A copy of the questionnaire is included in the appendix of this thesis as well (Appendix F)

CHAPTER 4: RESULTS

This chapter presents the outcomes from the study on the impact of air pollution on Acute Coronary Syndrome (ACS) patients, considering ACS hospitalization rates, ACS mortality rates, and mortality risk. Alongside, it also presents the results of the web-based prototype development, aiming at the effective use and visualization of these models by medical personnel. The results are organized into three main parts.

Section 4.1 presents the outcomes of the regression models, developed to predict hospitalization and mortality rates related to ACS. Section 4.2 presents the results of the classification models designed to predict mortality risk. Finally, section 4.3 of this chapter introduces the web system development prototype.

4.1 Regression Model Result

In this study, machine learning (ML) models were constructed to predict the rate hospitalization and mortality in ACS patients. Given the temporal nature of air pollution effects on health outcomes, the models were developed at four distinct time lags. These time lag phases were implemented to control potential delayed impacts and to provide deeper understanding of the association between air pollution and ACS outcomes. The models were designed to predict these two outcomes based on four key air quality parameters: Nitrogen Oxides (NOx), Sulfur Dioxide (SO_2), Ozone (O_3), and Particulate Matter 10 (PM10) on four varying timeframes of air pollution exposure, referred to as 'time lags'. Based on our preliminary study, time lag 00 demonstrated the best performance for ACS hospitalization and ACS mortality for all ML models. In addition, time lag 00 allows for easier integration and the air quality readings are easier to obtain. Therefore, this study focuses on presenting the results for time lag 00.

The time lag 00 accounts for daily air pollution exposure, time lag 03 accounts for three consecutive days of exposure, time lag 07 represents the average weekly exposure, and time lag 30 stands for the average monthly exposure. Where time lag 00 and time lag 03 are considered as short-term, and time lag 07 and time lag 30 as long-term.

Section 4.1.1 describes the baseline characteristics of the input and output variables across the four-time lag phases, evaluate the overall performance of the models, and present the importance of each feature using SHAP summary plots. The primary findings and implications of the regression analysis will be highlighted, providing insights into the significant effects of air quality on ACS patient outcomes.

4.1.1 Baseline Characteristics

The dataset used for the regression model comprised several attributes related to air quality, including NOx, SO_2, O_3, and PM10. Each attribute was analyzed for its potential association with the ACS hospitalization rate and ACS mortality rate among ACS patients.

A significant test was performed for each attribute against the ACS hospitalization rate and ACS mortality rate. The results indicated a high level of statistical significance for most of the air quality attributes, as evidenced by their p-values being less than 0.001. However, an exception was noted for PM10 in correlation with the admission rate, where the p-values for different lag times demonstrated no significant association. Specifically, at lag 00, the p-value was 0.096; at lag 03, the p-value was 0.056; at lag 07, the p-value was 0.336; and at lag 30, the p-value was 0.230. The p-values for PM10 at all lags (lag 00, lag 03, lag 07, and lag 30) are all above 0.05. This implies that the correlation between PM10 and the ACS hospitalization rate at these lag times is not statistically significant.

Analysis of the data reveals a mean daily admission rate of 1.22, with a standard deviation of 2.07. For the mean of daily ACS mortality rate is 0.08, with a standard deviation of 0.304. Table 4.1 provides an overview of the dataset used for the regression model, outlining the range and units of each attribute. In addition, the table summarizes the p-values for each attribute, thus showcasing their statistical significance in relation to both the admission rate and mortality rate of ACS patients.

Table 4.1: Baseline characteristics for air quality readings and hospitalization rate and mortality rate of ACS patients.

Variables	Attributes	Time lag 00	Time lag 03	Time lag 07	Time lag 30
		Value	Value	Value	Value
N	Total	57693	57692	8372	1904
Nitrogen Oxides (ppb)	Mean	89.81	89.81	89.77	89.81
	Std Dev	86.13	86.13	84.69	84.31
	Range	0 – 209.22	0 – 209.22	0 – 152.25	0 – 134.02
	p-value	**<0.001**	**<0.001**	**<0.001**	**<0.001**
Sulphur Dioxide (ppb)	Mean	77.49	77.49	77.44	77.78
	Std Dev	76.68	76.68	75.62	75.09
	Range	0 – 192.05	0 – 192.05	0 – 139.13	0 – 119.88
	p-value	**<0.001**	**<0.001**	**<0.001**	**<0.001**
Ozone (ppb)	Mean	85.03	85.03	84.99	85.03
	Std Dev	77.47	77.47	84.99	85.03
	Range	0 - 148.71	0 - 148.71	0 – 114.31	0 – 107.11
	p-value	**<0.001**	**<0.001**	**<0.001**	**<0.001**
Particulate Matter 10 ($\mu g/m^3$)	Mean	48.58	48.58	48.47	48.53
	Std Dev	23.64	23.64	20.29	16.66
	Range	0 – 515.0	0 – 515.0	0 – 285.14	0.54 – 169.77
	p-value	0.096	0.056	0.336	0.230
ACS Hospitalization Rate	Mean	1.22	1.22	8.43	48.0

Table 4.1, continued.

Variables	Attributes	Time lag 00	Time lag 03	Time lag 07	Time lag 30
		Value	Value	Value	Value
ACS Hospitalization Rate	Std Dev	2.07	2.07	11.49	48.0
	Range	0 - 26	0 - 26	0 - 96	0 – 332
ACS Mortality Rate	Mean	0.08	0.08	0.57	2.5
	Std Dev	0.304	0.304	1.026	3.384
	Range	0 – 4	0 – 4	0 – 8	0 – 24

4.1.2 Regression Models Performance

4.1.2.1 Model Performance

In the development of predictive models for hospitalization and mortality rates among ACS patients, this study utilized five distinct ML algorithms: Linear Regression, Support Vector Machine (SVM), XGBoost, Random Forest (RF), and an ensemble learning (EL) method with Generalized Linear Model (GLM) as the meta-learner. Each model was evaluated based on its Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), with the lower values indicating greater predictive accuracy.

The time lag 00 model demonstrated better performance matrices used in this study, presenting the lowest RMSE and MAE for both ACS hospitalization and ACS mortality rates among all the time lag phases assessed.

Table 4.2 and Table 4.3 below present a comparison of ML algorithms across four different time lags to predict the hospitalization and mortality rates for ACS patients. The RF model demonstrated the highest performance in predicting ACS patient hospitalization rates, attaining the lowest RMSE of 1.701 and MAE of 1.115 in time lag 00. Conversely, for predicting the ACS mortality rate, XGBoost provided the most accurate results, achieving the lowest RMSE of 0.440 and MAE of 0.194.

Table 4.2: Performance metrics (RMSE and MAE) of ML algorithms for predicting ACS patients' hospitalization rate across different time lags.

Time lag	**00**		**03**		**07**		**30**	
Performance Metrics	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**
Linear Regression	1.982	1.292	1.922	3.692	10.14	102.81	38.91	1514.2
Support Vector Machine (Linear)	2.089	1.213	2.003	4.013	10.53	110.99	41.59	1729.4
Random Forest	**1.701**	**1.115**	1.936	3.751	9.243	85.43	33.93	1151.3
XGBoost	1.846	1.202	1.888	3.566	9.169	84.07	35.86	1285.6
ENSEMBLE (GLM)	1.922	3.694	1.922	3.694	9.819	96.40	37.41	1399.2

Table 4.3: Performance metrics (RMSE and MAE) of ML algorithms for predicting ACS patients' mortality rate across different time lags.

Time lag	**00**		**03**		**07**		**30**	
Performance Metrics	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**
Linear Regression	0.443	0.197	0.444	0.197	0.995	0.990	3.101	9.617
Support Vector Machine (Linear)	0.461	0.213	0.461	0.213	1.125	1.266	3.354	11.246
Random Forest	0.442	0.195	0.444	0.197	0.979	0.959	2.724	7.419
XGBoost	**0.440**	**0.194**	0.447	0.200	0.964	0.929	2.841	8.073
ENSEMBLE (GLM)	0.444	0.197	0.444	0.197	0.989	0.978	2.981	8.889

Figure 4.1 and Figure 4.2 describe the relative importance of features associated for predicting hospitalization rates and mortality rates in ACS patients, respectively, as determined by ML models in this study.

NOx and O_3 consistently ranked as the most influential factors in all of ML models. Except in the RF model, where O_3 and SO_2 were ranked the highest for predicting ACS mortality rates. Furthermore, PM10 exhibited minimal impact across all ML models.

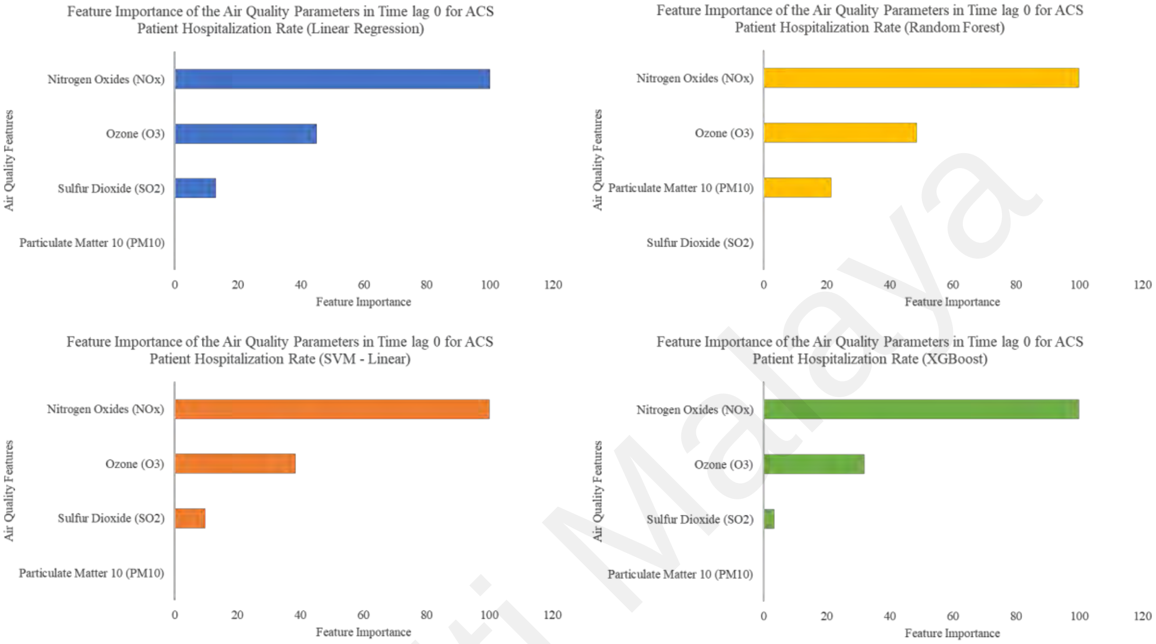


Figure 4.1: Relative feature importance for predicting hospitalization rate in ACS patients using different ML models at time lag 00.

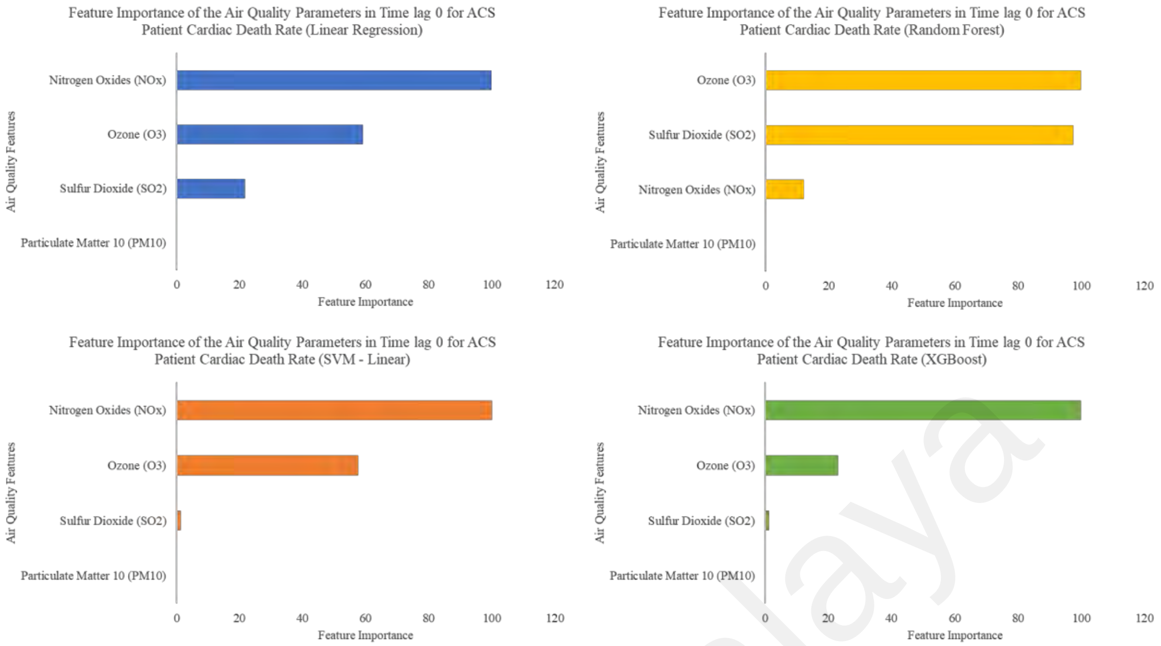


Figure 4.2: Relative features importance for predicting mortality rate in ACS Patients using different ML models at time lag 00.

Figure 4.3 offers a boxplot of the distribution of actual versus predicted hospitalization rates for ACS patients for time lag 00, as determined by each ML model using air quality readings at time lag 00. Similarly, Figure 4.4 provides a visual comparison of the actual and predicted mortality rates in ACS patients, based on data from the various ML models at time lag 00.

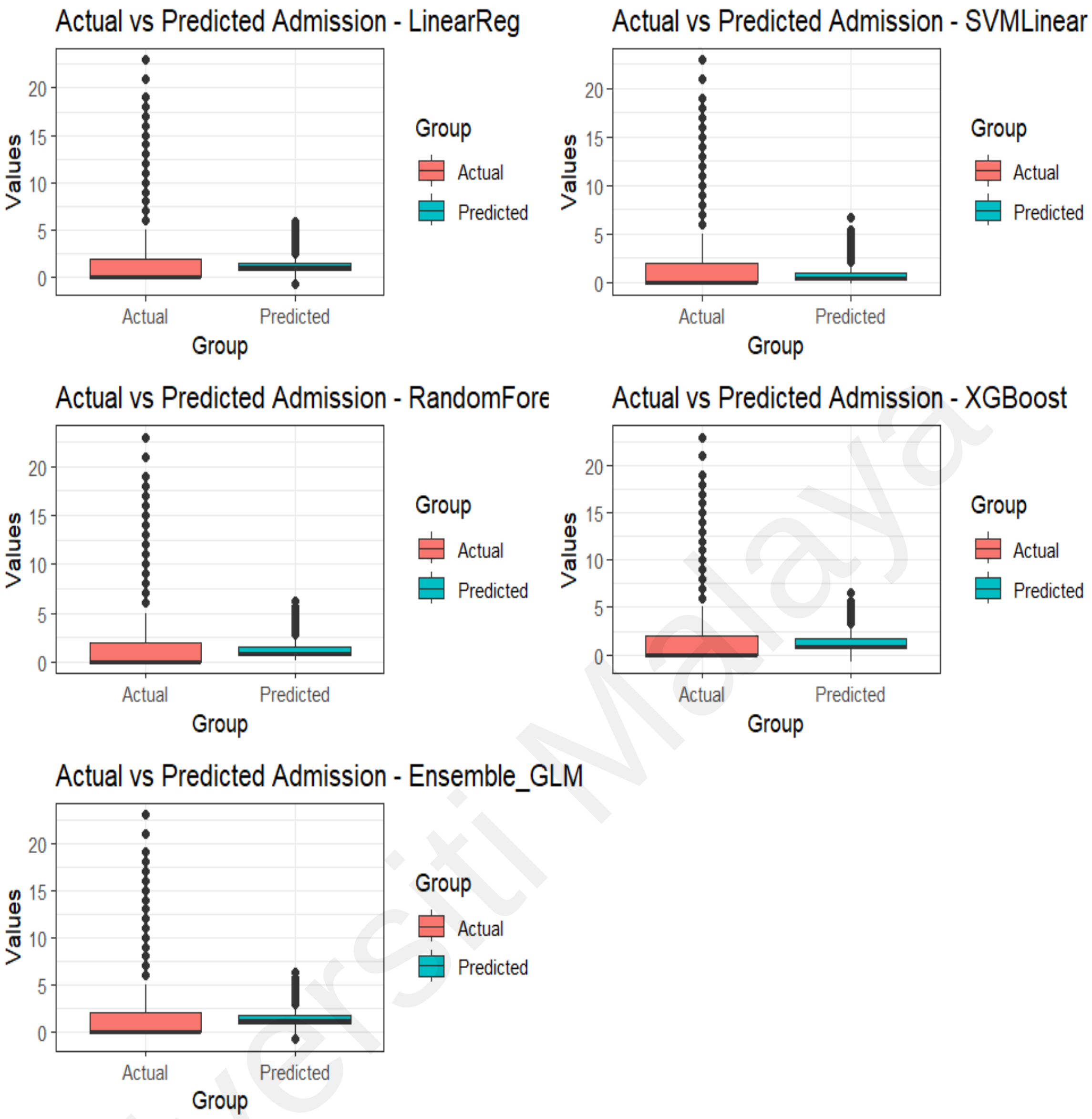


Figure 4.3: Boxplot illustrates the distribution of actual versus predicted hospitalization rates for ACS patients, derived from various ML models using time lag 00 air pollution data.

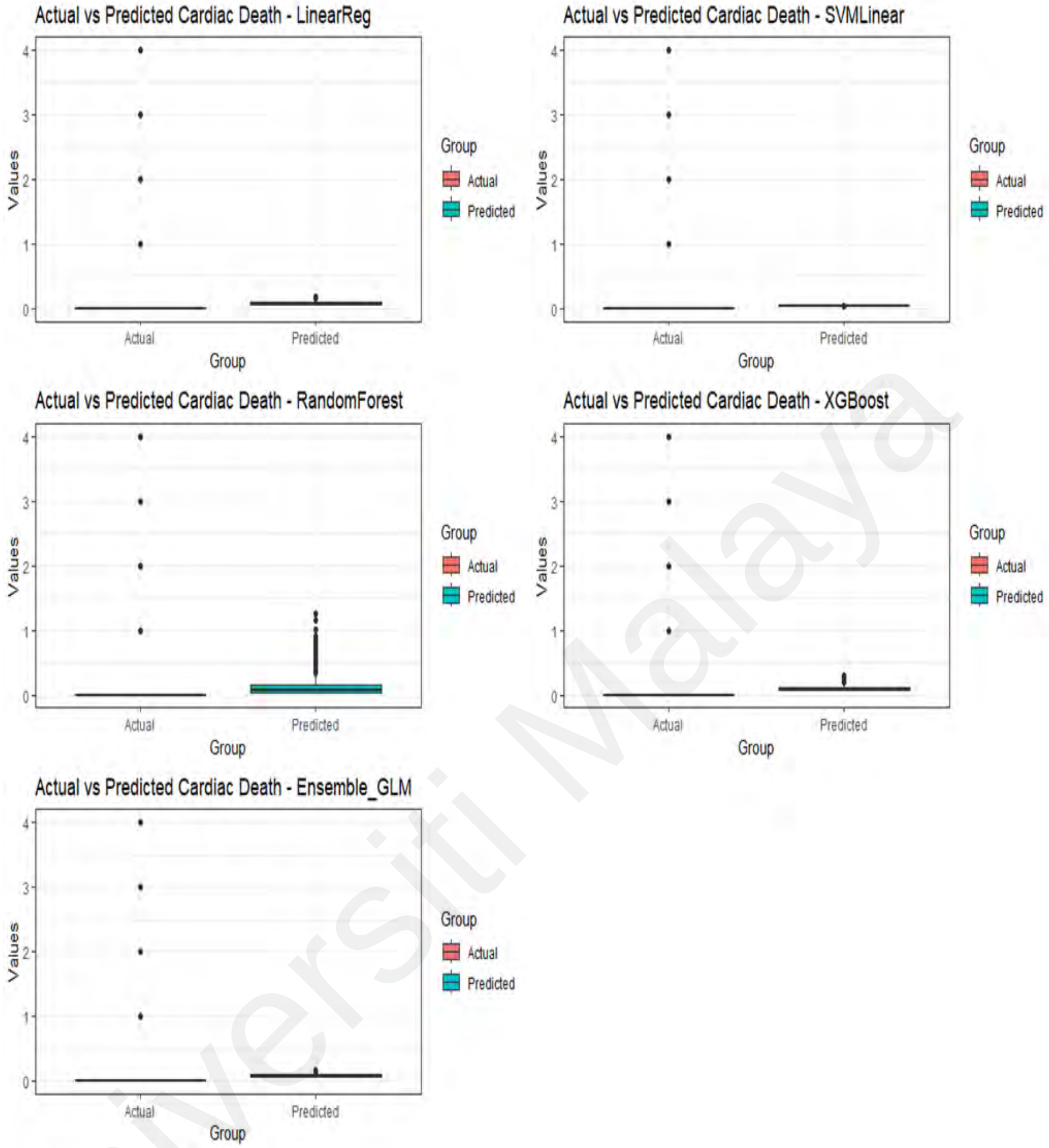


Figure 4.4: Boxplot illustrates the distribution of actual versus predicted mortality rates in ACS patients, as determined by different ML models using time lag 00 air pollution data.

The boxplot below, Figure 4.5 and Figure 4.6, provide a clearer picture of the best model's performance. Specifically, Figure 4.6 presents a boxplot of actual and predicted hospitalization rates for ACS patients based on the RF model. Similarly, Figure 4.7 showcases the XGBoost model's ability to predict the mortality rates of ACS patients at time lag 0, as relative with the actual rates.

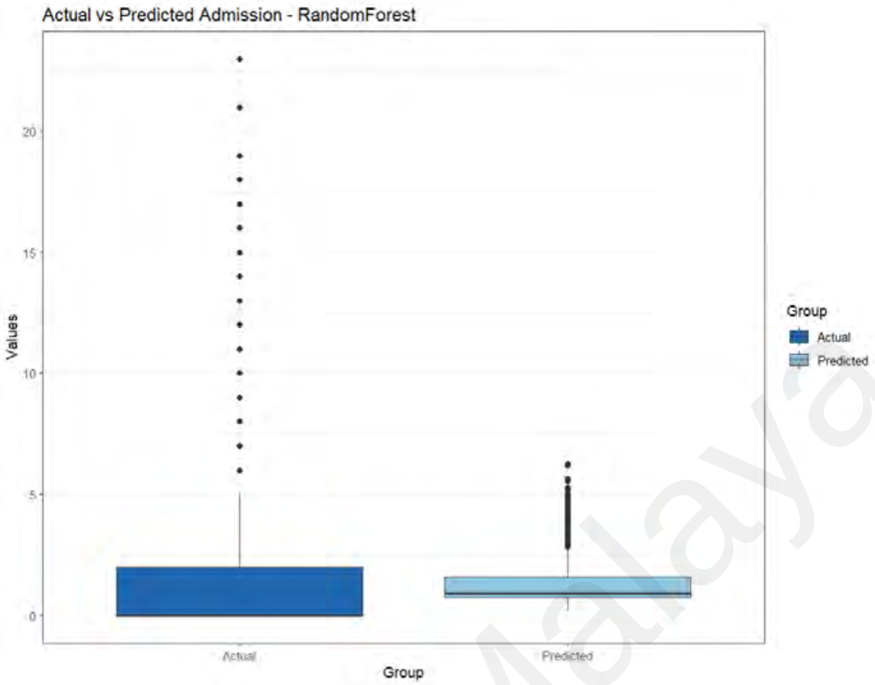


Figure 4.5: Boxplot depicting the actual versus predicted hospitalization rates of ACS patients at time lag 0, as predicted by the RF model.

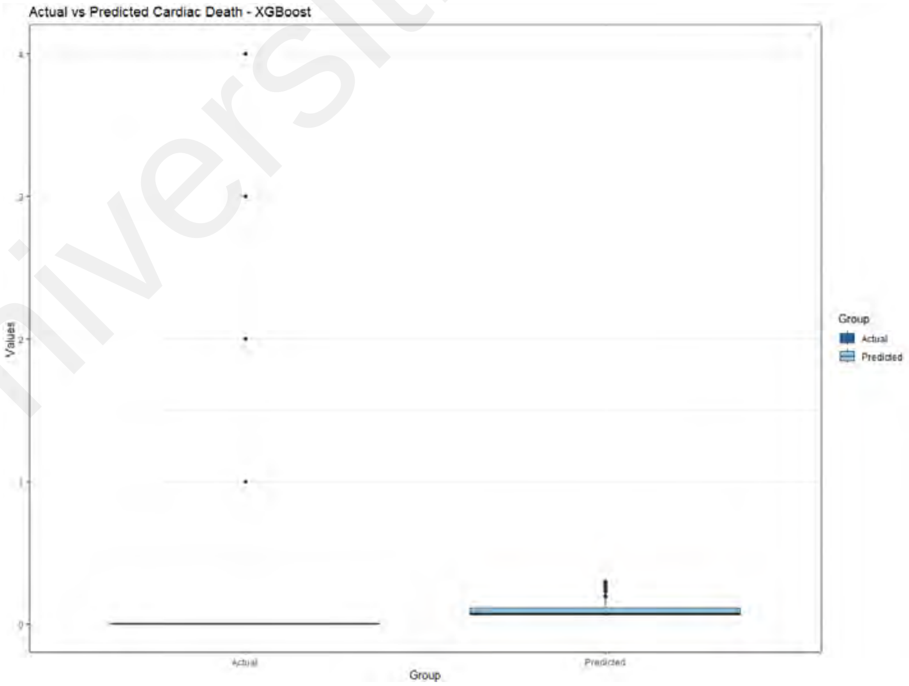


Figure 4.6: Boxplot illustrating the actual versus predicted mortality rates of ACS patients at time lag 0, as predicted by the XGBoost model.

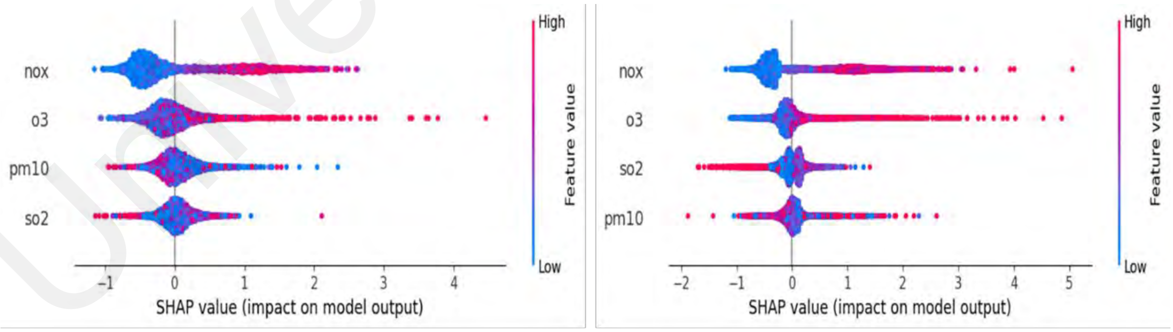
4.1.2.2 SHAP Analysis

Shapley Additive Explanations (SHAP) was implemented after model training for ML model interpretation.

Figure 4.7 presents the SHAP summary plot which visualizes the top features of the RF model's output which is the hospitalization rate of ACS patients. Each instance is represented by a single dot on each feature row, the x-axis shows the SHAP value, and the y-axis shows the feature name.

In the context of predicting ACS hospitalization rates, the SHAP summary plot (Figure 4.7(a)) shows that NOx and O_3 significantly impact the model's predictions. Higher values of NOx and O_3, indicated in red, contribute to an increase in hospitalization rate predictions, implying a positive effect on the model.

The same analysis was applied to the prediction of ACS mortality rates, and the results were similar. The SHAP summary plot (Figure 4.7(b)) indicates that NOx and O_3 are the most influential features, while compared against SO_2 and PM10.



(a) (b)

Figure 4.7: SHAP summary plot for the regression ML model. (a) The RF model predicting the hospitalization rate and (b) the XGBoost model predicting mortality rate of ACS patients, illustrating the impact of each feature on the model's predictions.

4.2 Classification Model Result

This section presents the results of the developed classification models to predict in-hospital and emergency mortality risk of ACS patients in the presence of air pollution. The classification models are developed incorporating features related to in-hospital and emergency settings and include daily air quality measurements for ACS patients.

The in-hospital features selected were adapted from our previous study, which used a SVM model with a variable importance sequential backward elimination method to identify the top 14 variables for predicting mortality risk in the same ACS cohort (Kasim, et al., 2022). These features were combined with air quality variables, underlining the study's focus on the impact of air pollution on ACS patients. Conversely, the emergency features consist of a reduced set of variables that are easily accessible in Malaysia hospital emergency settings, without the need for extensive testing or patient history.

The results include the baseline characteristics of our patient cohort and summary of model performance matrices, including detailed breakdown of model performance, Receiver Operating Characteristics (ROC) curves, and SHAP values.

In addition, the classification model based on ML is compared with the Thrombolysis in Myocardial Infarction (TIMI) Score, using the Net Reclassification Improvement (NRI) and the risk cut-off point that differentiates between low and high-risk patients.

4.2.1 Baseline Characteristic

Table 4.4 presents the summary statistics of 14,145 in-hospital and 22,466 emergency ACS patients, selected from the complete dataset. The in-hospital dataset comprises 18 features while the emergency dataset includes 12, chosen particularly for their relevance to emergency situations.

In the ACS patient dataset, the only demographic feature considered was age. This feature showed that the average age of ACS patients was 59 years (SD = 39), a statistic consistent across both in-hospital and emergency cases.

A statistical difference ($p<0.001$) was observed between survivors and non-survivors, however, between the in-hospital and emergency features, the survival determinant factors differed slightly. In in-hospital selected variables, age, heart rate, Killip class, fasting blood glucose, HDLC, LDLC, usage of statins, oral hypoglycemic agents, anti-arrhythmic agents, and exposure to NOx and O_3 showed a statistically significant association ($p<0.001$).

In the emergency selected variables identified a different group of influential factors, these included age, heart rate, Killip class, ECG abnormalities, use of statins, oral hypoglycemic agents, anti-arrhythmic agents, and exposure to NOx and O_3 where all variables have p-values <0.001. Particularly, ECG abnormalities, which were not significant in the in-hospital dataset, emerged as an important variable in the emergency dataset.

The mortality rates were 6.1% and 8.9% for in-hospital and emergency cases respectively, indicating the need for data balancing for accurate model development. Among the ACS patients, 58.47% of in-hospital and 55.38% of emergency cases were diagnosed with STEMI. NSTEMI and Unstable Angina (UA) together accounted for the remaining cases in both datasets, and both groups were compared against the TIMI risk score.

Table 4.4: Summary statistics for in-hospital selected variables and emergency variables.

Variables	**Features**	**In-Hospital Selected Variables**				**Emergency Variables**			
		All cases (14145)	Survivors (13287)	Non-survivors (858)	*p*-value	All cases (22466)	Survivors (20457)	Non-survivors (2009)	*p*-value
ACS Stratum	STEMI	8271 (58.5%)	7659 (57.6%)	612 (71.3%)	**<0.001**	12441 (55.4%)	11020 (53.9%)	1421 (70.7%)	**<0.001**
	NSTEMI	3460 (24.5%)	3244 (24.4%)	216 (25.2%)		5878 (26.2%)	5373 (26.3%)	505 (25.1%)	
	UA	2414 (17.1%)	2384 (17.9%)	30 (3.5%)		4147 (18.5%)	4046 (19.8%)	83 (4.1%)	
Age*		20.9 ±96.6	20.9±96.6	23.2 ±92.2	**<0.001**	20.9 ±97.6	20.9±97.6	21.1 ±96.9	**<0.001**
Heart Rate*		22±200	27±200	22±182	**<0.001**	20±200	20±200	22±194	**<0.001**
Chronic Angina (<2 weeks)		9610 (67.9%)	9031 (68.0%)	579 (67.5%)	0.767	15121 (67.3%)	13799 (67.5%)	1322 (65.8%)	0.133
Killip class*	I:	9767 (69.0%)	9561 (72.0%)	206 (24.0%)	**<0.001**	15234 (67.8%)	14722 (72.0%)	512 (25.5%)	**<0.001**
	II:	2712 (19.2%)	2520 (19.0%)	192 (22.4%)		4344 (19.3%)	3887 (19.0%)	457 (22.7%)	
	III:	659 (4.7%)	545 (4.1%)	114 (13.3%)		1080 (4.8%)	851 (4.2%)	229 (11.4%)	
	IV:	1007 (7.1%)	661 (5.0%)	346 (40.03%)		1808 (8.0%)	997 (4.9%)	811 (40.4%)	
ECG Abnormalities**		3967 (28.0%)	3688 (27.8%)	279 (32.5%)	0.003	6068 (27.0%)	5411 (26.5%)	657 (32.7%)	**<0.001**
HDL*		0.50 ±4.94	0.50±4.94	0.50 ±3.00	**<0.001**				
LDL*		0.50 ±18.0	0.60±18.0	0.50 ±9.44	**<0.001**				
Fasting Blood Glucose*		3.00 ±49.0	3.00±49.0	3.00 ±46.4	**<0.001**				
Cardiac Catheterization		5166 (36.5%)	4878 (36.7%)	288 (33.6%)	0.064				
Coronary Artery Bypass Graft (CABG)		124 (0.9%)	114 (0.9%)	10 (1.2%)	0.349				

Table 4.4, continued.

Variables	**Features**	**In-Hospital Selected Variables**				**Emergency Variables**			
		All cases (14145)	Survivors (13287)	Non-survivors (858)	*p*-value	All cases (22466)	Survivors (20457)	Non-survivors (2009)	*p*-value
Statin*		13278 (93.9%)	12533 (94.3%)	745 (86.8%)	**<0.001**	20597 (91.7%)	19038 (93.1%)	1559 (77.6%)	**<0.001**
Other Lipid Lowering Agent		434 (3.1%)	422 (3.2%)	12 (1.4%)	0.003	774 (3.4%)	726 (3.5%)	48 (2.4%)	0.007
Oral Hypoglycemic Agent*		3424 (24.2%)	3340 (25.1%)	84 (9.8%)	**<0.001**	5296 (23.6%)	5115 (25.0%)	181 (9.0%)	**<0.001**
Antiarrhythmic agent*		680 (4.8%)	570 (4.3%)	110 (12.8%)	**<0.001**	272 (1.2%)	997 (4.9%)	1269 (63.2%)	**<0.001**
Nitrogen Oxides*		0±209.22	0±137.74	0 ±187.87	**<0.001**	0±137.74	0±137.74	0 ±159.39	**<0.001**
Sulfur Dioxide		0±192.05	0±207.03	0 ±211.81	0.024	0±178.13	0±178.13	0 ±211.81	0.062
Ozone*		0±148.71	0±129.91	0 ±124.71	**<0.001**	0±129.91	0±129.91	0 ±124.71	**<0.001**
Particulate Matter 10		0±390	0±390	0 ±322	0.643	0±390	0±390	0±372	0.016

The asterisk (*) indicated that the variable difference between the survivor and non-survivor groups is statistically significant (p-value <0.001). Particularly, the double asterisk (**) indicated for ECG Abnormalities, indicating that this variable was found statistically significant within the emergency-selected variables dataset. The significant values are given in bold. HDL: High Density Lipoprotein; LDL: Low Density Lipoprotein.

4.2.2 Classification Models Performance

4.2.2.1 Algorithm Performance Evaluation

Table 4.5 illustrates the classification model performances developed in this study using the selected in-hospital and emergency features based on the remaining 30% testing dataset. The results show that ML algorithms and EL approach significantly outperformed the TIMI risk scores in predicting both STEMI and NSTEMI outcomes in the presence of air pollution.

In the context of in-hospital selected features, the RF model demonstrated the high predictive performance of achieving an AUC of 0.843 (95% CI: 0.813 - 0.873) (p-value < 0.001). For emergency selected features, XGBoost algorithm yielded the highest AUC, achieving a score of 0.845 (95% CI: 0.828 - 0.862) (p-value < 0.001).

The TIMI risk score showed a comparatively lower performance for both feature sets. In the in-hospital settings, the AUC for TIMI was found to be 0.791 and 0.565 for STEMI and NSTEMI respectively. In emergency settings, the AUC for TIMI was lower, with scores of 0.797 and 0.583 for STEMI and NSTEMI respectively. It is noticeable that TIMI predicting the risk of mortality of STEMI patients are still within the acceptable range, whereas the AUC for TIMI NSTEMI patients performed poorly as compared to the other predictive models.

Detailed performance evaluation of the best ML model against TIMI risk score for in-hospital and emergency selected features are presented in Table 4.6.

Table 4.5: The AUC of ML models and TIMI risk score for in-hospital selected features and emergency selected features based on 30% testing dataset.

Predictive Models	**The area under the ROC Curve (95% CI)**	
	In-Hospital Selected Features	**Emergency Selected Features**
Logistic Regression	0.834 (0.803 - 0.865)	0.842 (0.825 - 0.859)
SVM (Linear)	0.833 (0.803 - 0.864)	0.842 (0.825 - 0.86)
Random Forest	**0.843 (0.813 - 0.873)**	0.843 (0.826 - 0.86)
Naïve Bayes	0.838 (0.807 - 0.869)	0.834 (0.816 - 0.852)
XGBoost	0.836 (0.804 - 0.868)	**0.845 (0.828 - 0.862)**
Ensemble (GLM)	0.842 (0.812 - 0.873)	0.844 (0.828 - 0.862)
TIMI (STEMI)	0.791 (0.757 - 0.825)	0.797 (0.774 - 0.82)
TIMI (NSTEMI)	0.565 (0.505 - 0.625)	0.583 (0.543 - 0.622)

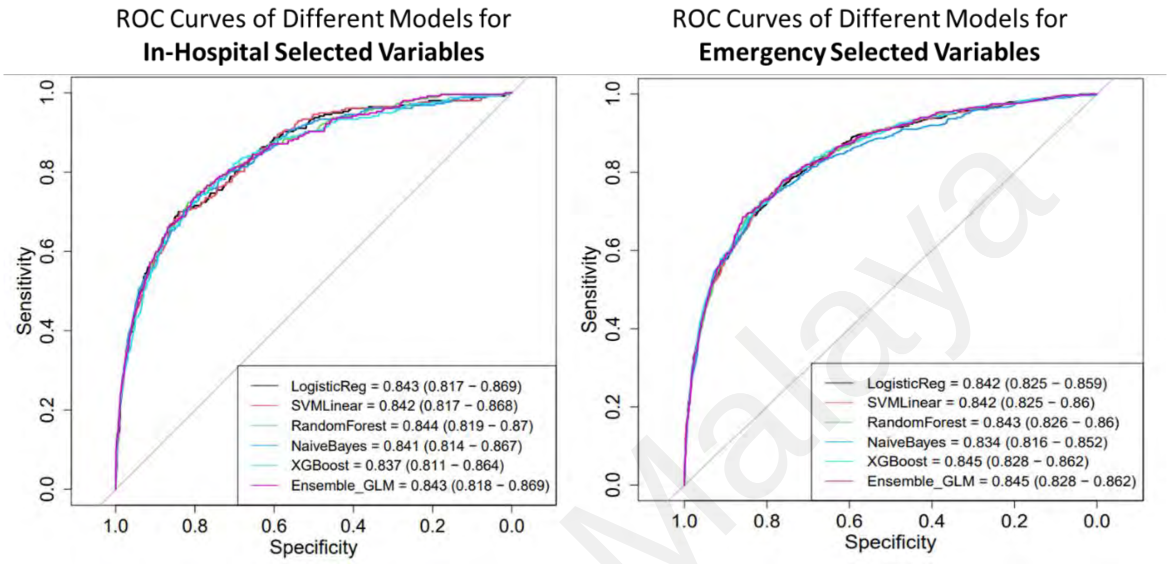
Table 4.6: Detailed performance metrics of ML model for in-hospital and emergency selected features for ACS patients.

No.	Predictive Models	Accuracy	Sensitivity	Specificity	PPV	NPV	McNemar Test	Balanced Accuracy	Precision Recall Accuracy
In- Hospital Selected Features Dataset									
1.	Logistic Regression	0.82 (0.804 - 0.834)	0.684	0.829	0.227	0.973	0	0.757	0.354
2.	SVM (Linear)	0.821 (0.805 - 0.836)	0.673	0.832	0.226	0.972	0	0.752	0.347
3.	**Random Forest**	0.849 (0.834 - 0.863)	0.632	0.865	0.255	0.970	0	0.748	0.372
4.	Naïve Bayes	0.875 (0.862 - 0.888)	0.538	0.900	0.283	0.964	0	0.719	0.357
5.	XGBoost	0.847 (0.832 - 0.861)	0.667	0.860	0.259	0.972	0	0.763	0.364
6.	Ensemble (GLM)	0.846 (0.832 - 0.86)	0.643	0.861	0.253	0.971	0	0.752	0.374
7.	TIMI (STEMI)	0.807 (0.791 - 0.822)	0.585	0.823	0.195	0.964	0	0.704	0.240
8.	TIMI (NSTEMI)	0.936 (0.923 - 0.947)	0.035	0.982	0.094	0.951	0	0.509	0.061

Table 4.6, continued.

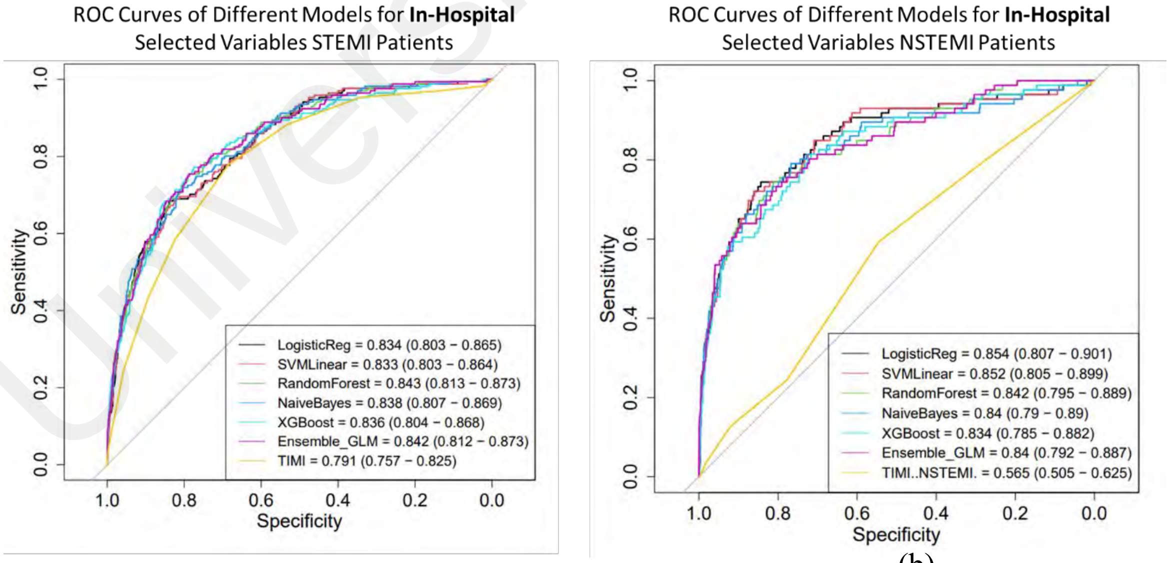
No.	Predictive Models	Accuracy	Sensitivity	Specificity	PPV	NPV	McNemar Test	Balanced Accuracy	Precision Recall Accuracy
Emergency Selected Features Dataset									
1.	LogisticReg	0.804 (0.794 - 0.813)	0.709	0.813	0.271	0.966	0	0.761	0.445
2.	SVMLinear	0.807 (0.798 - 0.817)	0.713	0.817	0.276	0.967	0	0.765	0.442
3.	RandomForest	0.809 (0.799 - 0.818)	0.724	0.817	0.279	0.968	0	0.771	0.434
4.	NaiveBayes	0.869 (0.86 - 0.877)	0.581	0.897	0.356	0.956	0	0.739	0.433
5.	**XGBoost**	0.813 (0.803 - 0.822)	0.704	0.824	0.281	0.966	0	0.764	0.451
6.	Ensemble_GLM	0.804 (0.794 - 0.813)	0.724	0.811	0.274	0.968	0	0.768	0.448
7.	TIMI (STEMI)	0.792 (0.778 - 0.805)	0.656	0.809	0.305	0.949	0	0.733	0.375
8.	TIMI (NSTEMI)	0.923 (0.913 - 0.932)	0.022	0.981	0.069	0.940	0	0.501	0.075

The ROC curve for the predictive models based on the testing dataset is shown in Figure 4.8. ROC curves for in-hospital and emergency mortality predictions, stratified by STEMI and NSTEMI, are presented in Figures 4.9 and 4.10, respectively.



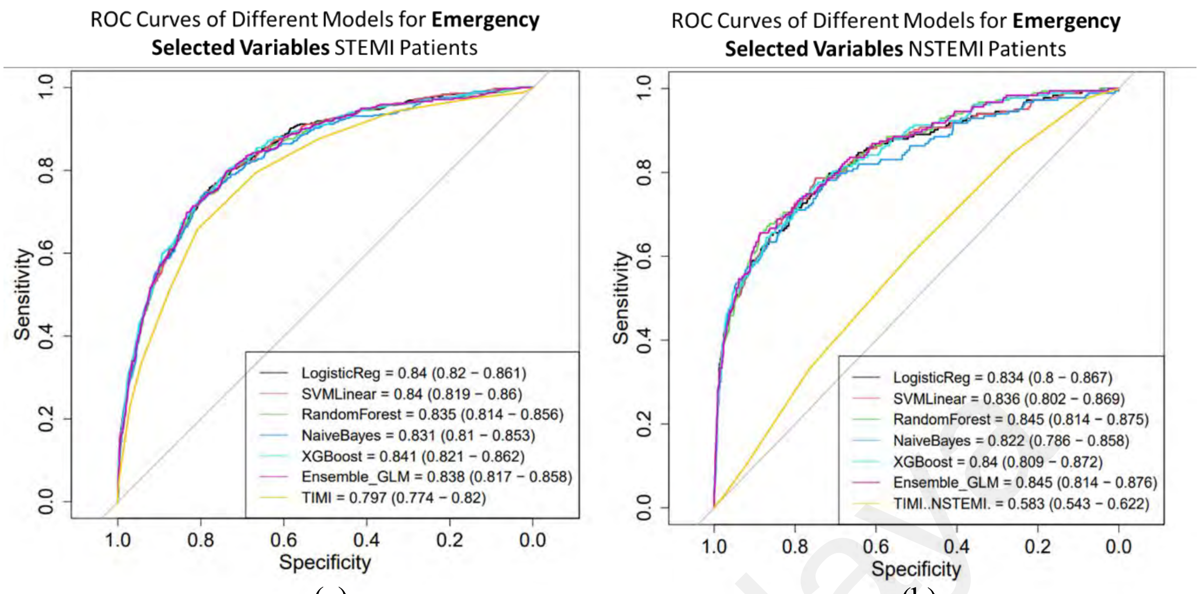
(a) (b)

Figure 4.8: ROC curve for ML and EL models in testing dataset for (a) in-hospital selected variables and (b) emergency selected variables.



(a) (b)

Figure 4.9: ROC Curves of ML models, EL model and TIMI for in-Hospital selected variables mortality prediction for (a) STEMI and (b) NSTEMI patients.



(a) (b)

Figure 4.10: ROC Curves of ML models, EL model and TIMI for emergency selected variables mortality prediction for (a) STEMI and (b) NSTEMI patients.

4.2.2.2 SHAP Analysis

Figures 4.11 and 4.12 display the SHAP summary plots for the RF with in-hospital selected features and XGBoost with emergency selected features, respectively. These plots offer a detailed view of feature importance, merging it with the effects of each feature on the testing dataset. The gradient color indicates the variable's initial value. In Booleans, it can contain two colors, but in numbers, it can contain the entire color spectrum. Each point corresponds to a row in the initial dataset. The color of the dots denotes the value of the feature (Blue: low value; Red: Higher blue). Features are well-organized depending on their importance during the interaction. The y-axis indicates the variable name in descending order of importance, with Killip classification having the highest importance in both models. On the x-axis indicates the SHAP value.

Considering the SHAP summary plot for the RF model (Figure 4.11) with in-hospital selected variables, features such as Killip Class, Fasting Blood Glucose (FBG), patient's age, heart rate, and usage of oral hypoglycemic agents are linked with higher negative effects on the outcome. This association suggests that an increase in these features correlates with an increase mortality risk. In addition, NOx was found to have the strongest association with mortality risk among in-hospital ACS patients.

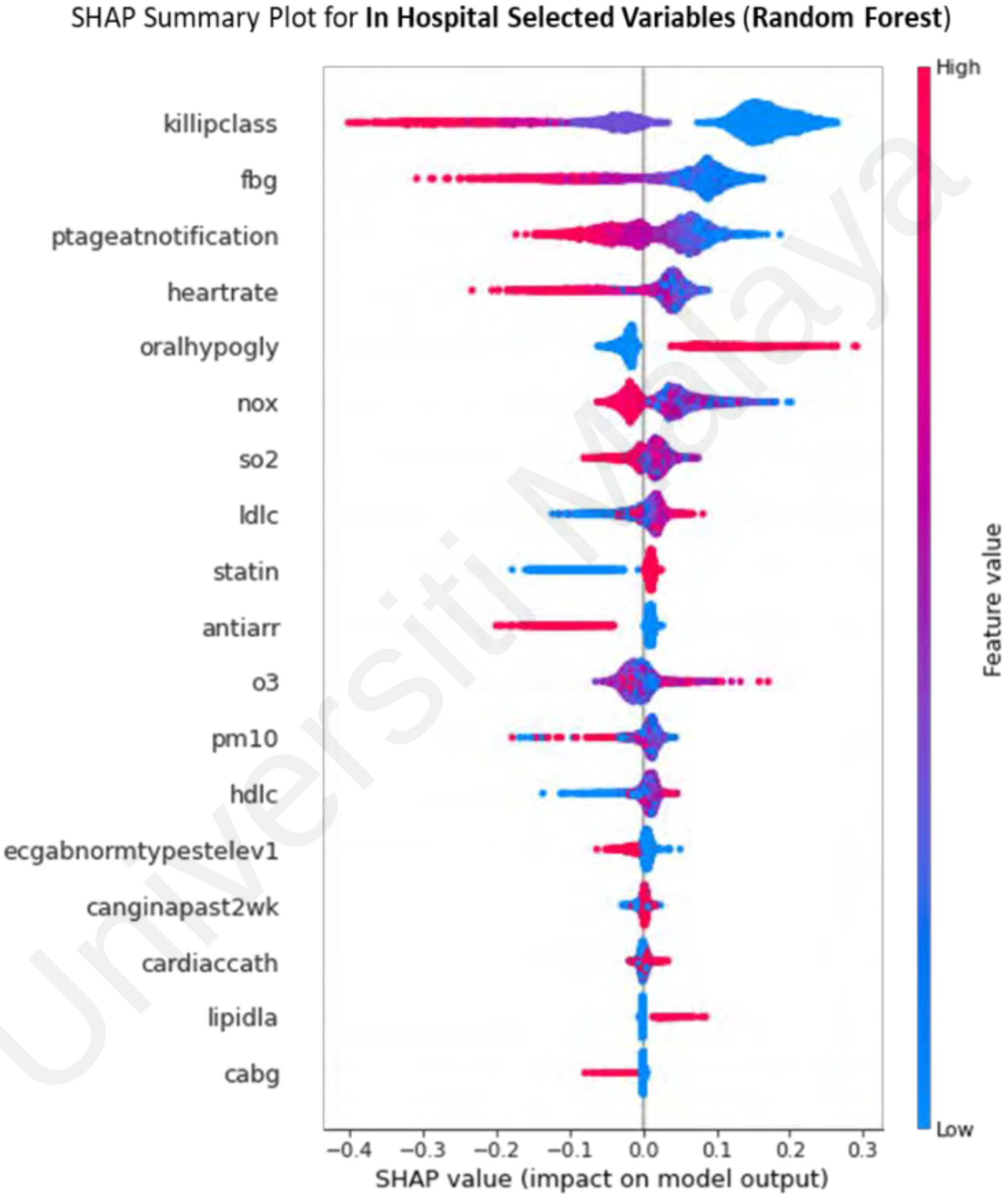


Figure 4.11: SHAP summary plot of RF model based on in-hospital selected features.

The XGBoost model SHAP summary plot for the emergency features dataset (Figure 4.12) revealed that the most significant features were Killip Class, patient's age, heart rate, intake of oral hypoglycemic agents, and statins. In the context of air pollutants, higher NOx values are associated with higher mortality risk. Overall, the SHAP summary plots provide a comprehensive understanding of the influence and importance of different variables in our ML models.

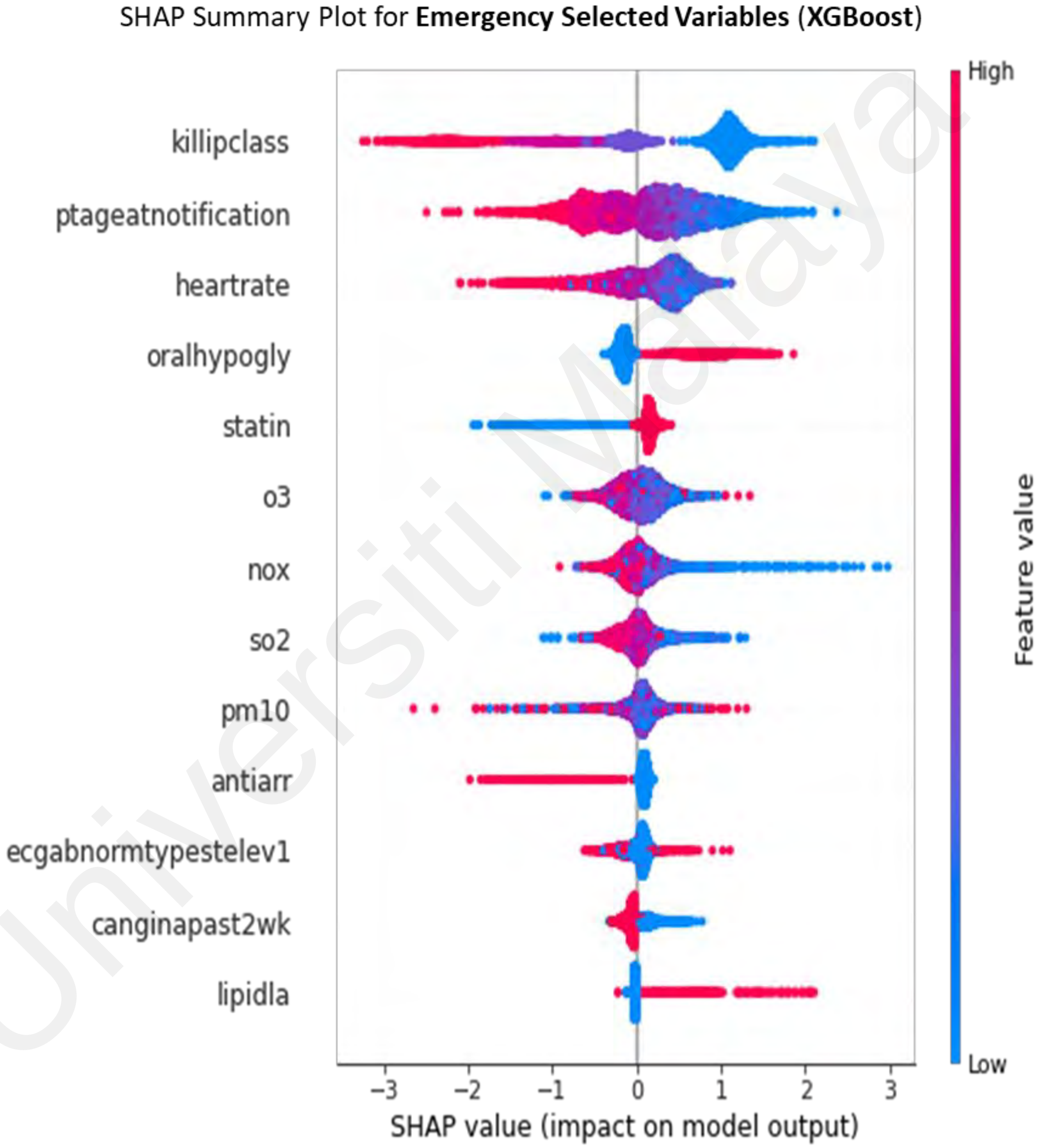


Figure 4.12: SHAP summary plot of XGBoost model based on emergency selected features.

4.2.2.3 Comparison of Machine Learning (ML) to Thrombolysis in Myocardial Infarction (TIMI) Risk Score to the Validation Dataset

The TIMI score for STEMI categories 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 categorizes 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 used in this study, the classification of patients into low- and high-risk categories was achieved based on the Receiver Operating Characteristic (ROC) curve approach, effectively measuring the trade-off between sensitivity and specificity across a series of cut-off points for model performance assessments (Kumar & Indrayan, 2011). Hence, the cut-off points between the low- and the high-risk patients for TIMI risk score and all the best ML models are presented in Figure 4.13 below.

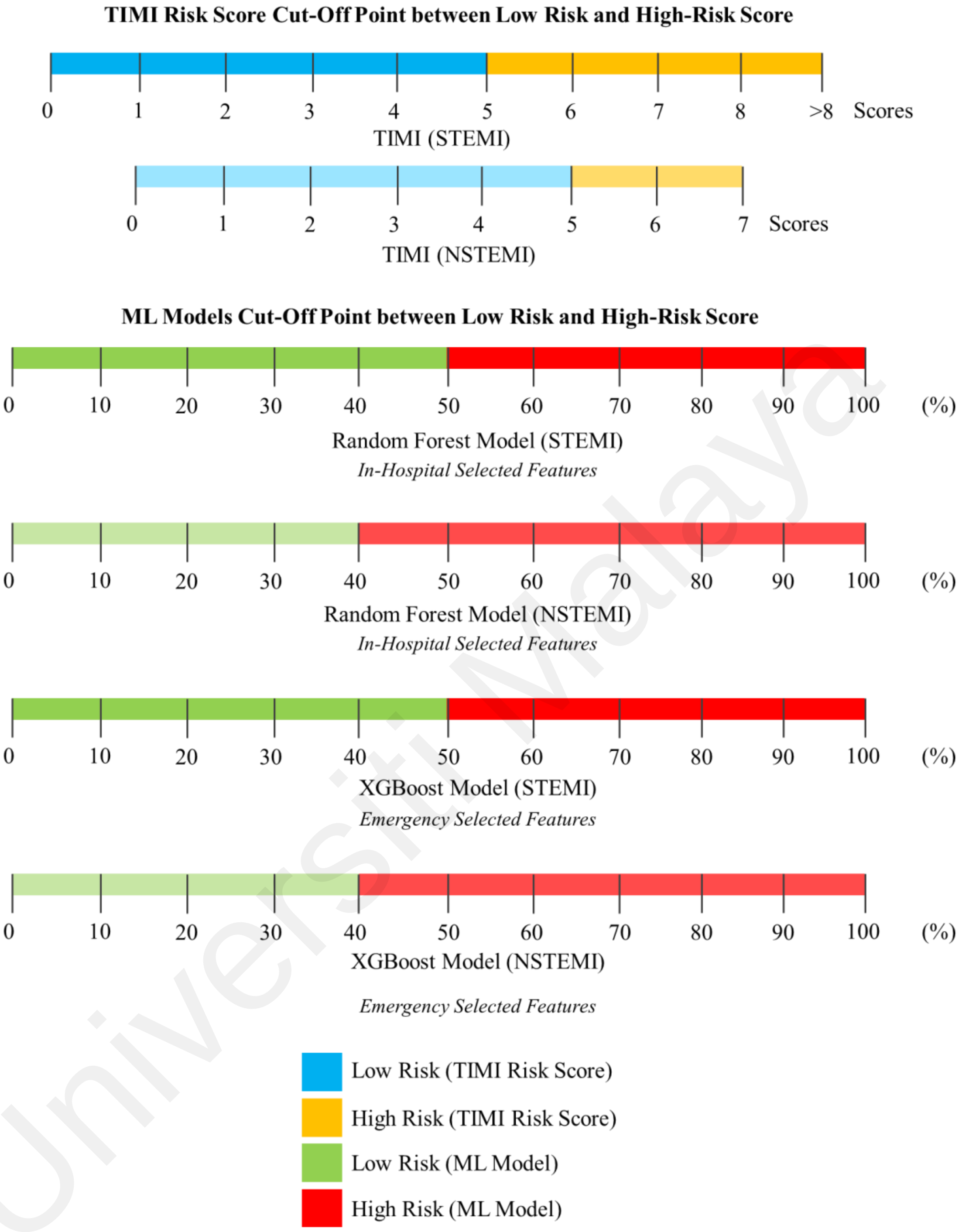
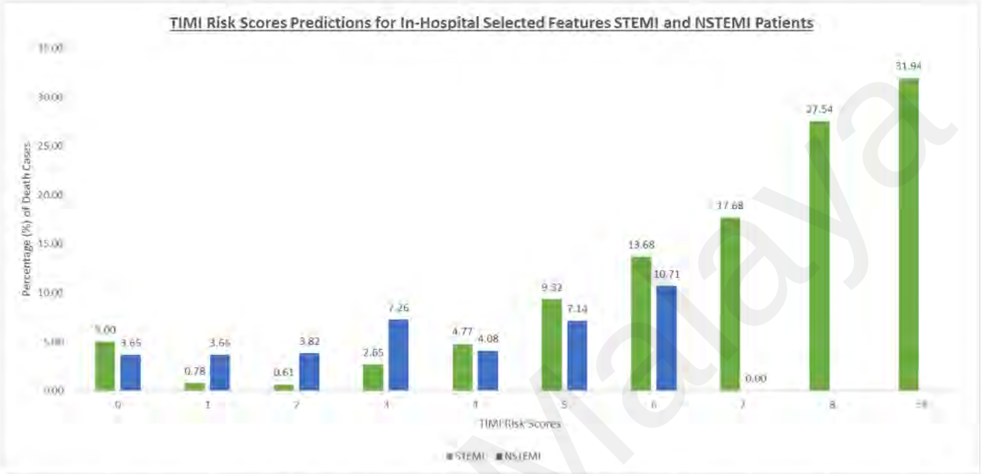

TIMI Risk Score Cut-Off Point between Low Risk and High-Risk Score
0 1 2 3 4 5 6 7 8 >8 Scores
TIMI (STEMI)
0 1 2 3 4 5 6 7 Scores
TIMI (NSTEMI)
ML Models Cut-Off Point between Low Risk and High-Risk Score
0 10 20 30 40 50 60 70 80 90 100 (%)
Random Forest Model (STEMI)
In-Hospital Selected Features
0 10 20 30 40 50 60 70 80 90 100 (%)
Random Forest Model (NSTEMI)
In-Hospital Selected Features
0 10 20 30 40 50 60 70 80 90 100 (%)
XGBoost Model (STEMI)
Emergency Selected Features
0 10 20 30 40 50 60 70 80 90 100 (%)
XGBoost Model (NSTEMI)
Emergency Selected Features
Low Risk (TIMI Risk Score)
High Risk (TIMI Risk Score)
Low Risk (ML Model)
High Risk (ML Model)

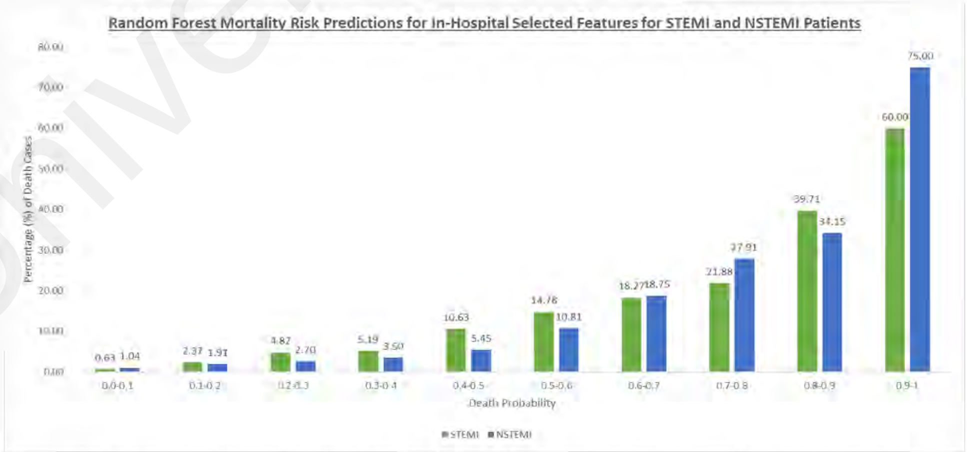
Figure 4.13: TIMI risk score and best performing ML models cut-off point between low-risk and high-risk group (Antman, et al., 2000; Morrow, et al., 2000; Kumar & Indrayan, 2011).

Figures 4.14 and 4.15 illustrate the comparison of the best ML model for (RF model) mortality risk against the TIMI risk score for both STEMI and NSTEMI. Similarly, in Figures 4.16 and 4.17 are for emergency selected features mortality rate. TIMI Risk Score for STEMI has a scale of 0–14 while TIMI Risk Score for NSTEMI has a scale of 0–7. We categorized ML score patients as low risk with the probability <50% and high-risk stratum as ≥50%. This is equivalent to TIMI low risk of score ≤5 and a high-risk score of > 5 for both STEMI and NSTEMI risk scores (Basra, et al., 2016; Kumar & Cannon, 2009).



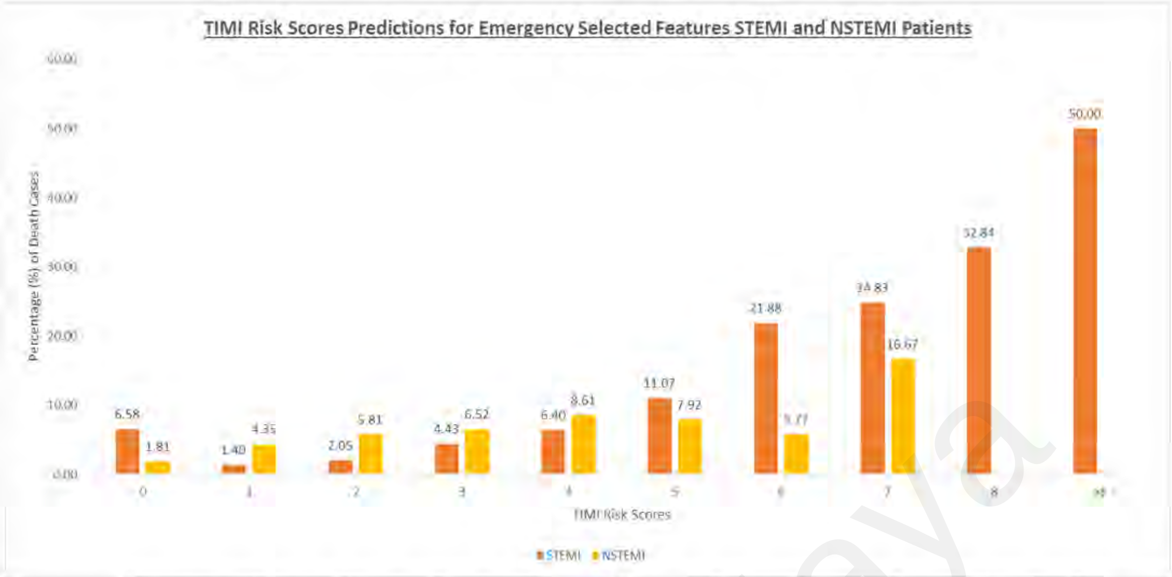
Total Number of Patients	STEMI	60	258	492	452	377	354	190	181	69	72
	NSTEMI	137	329	471	413	245	112	28	4		
Total Number of Dead Cases	STEMI	3	2	3	12	18	33	26	32	19	23
	NSTEMI	5	12	18	30	10	8	3	0		
Total Number of Alive Cases	STEMI	57	256	489	440	359	321	164	149	50	49
	NSTEMI	132	317	453	383	235	104	25	4		

Figure 4.14: Performance breakdown of the TIMI risk score for in-hospital selected variables mortality prediction for both STEMI and NSTEMI patients.



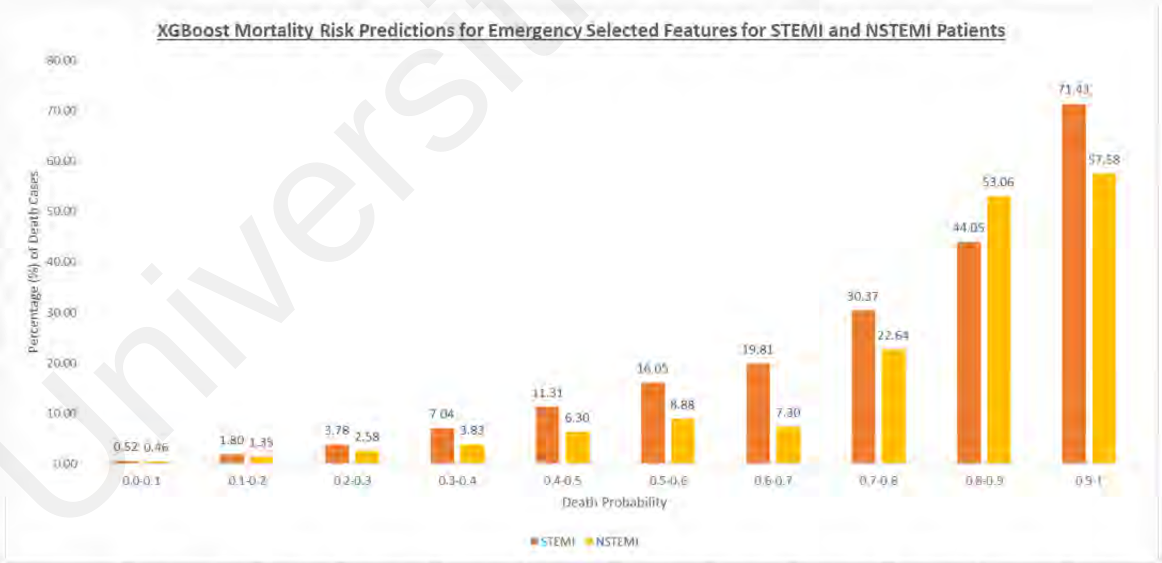
Total Number of Patients	STEMI	788	590	332	212	160	115	104	96	68	40
	NSTEMI	676	366	222	143	110	74	48	43	41	16
Total Number of Dead Cases	STEMI	5	14	16	11	17	17	19	21	27	24
	NSTEMI	7	7	6	5	6	8	9	12	14	12
Total Number of Alive Cases	STEMI	783	576	316	201	143	98	85	75	41	16
	NSTEMI	669	359	216	138	104	66	39	31	27	4

Figure 4.15: Performance breakdown of the ML model (RF model) for in-hospital selected variables mortality prediction for both STEMI and NSTEMI patients.



Total Number of Patients	STEMI	76	358	682	632	531	524	288	290	134	190
	NSTEMI	221	552	774	751	476	202	52	6		
Total Number of Dead Cases	STEMI	5	5	14	28	34	58	63	72	44	95
	NSTEMI	4	24	45	49	41	16	3	1		
Total Number of Alive Cases	STEMI	71	353	668	604	497	466	225	218	90	95
	NSTEMI	217	528	729	702	435	186	49	5		

Figure 4.16: Performance breakdown of the TIMI risk score for emergency selected variables mortality prediction for both STEMI and NSTEMI patients.



Total Number of Patients	STEMI	191	1058	740	426	327	243	212	191	227	91
	NSTEMI	219	963	621	418	270	169	137	106	98	33
Total Number of Dead Cases	STEMI	1	19	28	30	37	39	42	58	100	65
	NSTEMI	1	13	16	16	17	15	10	24	52	19
Total Number of Alive Cases	STEMI	190	1039	712	396	290	204	170	133	127	26
	NSTEMI	218	950	605	402	253	154	127	82	46	14

Figure 4.17: Performance breakdown of the ML model (XGBoost model) for emergency selected variables mortality prediction for both STEMI and NSTEMI patients.

In the context of in-hospital selected features, the RF model correctly classified 25.53% of STEMI patients and 18.37% of NSTEMI patients as high risk (risk probability greater than 50%), in comparison with TIMI score, it correctly classified 19.53% for STEMI patients and 9.38% for NSTEMI patients. The XGBoost model correctly classified 31.54% of STEMI patient and 16.85% for NSTEMI patient, where TIMI scores correctly classified 30.38% for STEMI patients and 6.01% for NSTEMI patients.

Table 4.7 tabulates the percentage of mortality in the patients with predicted low-risk (TIMI score: <5; ML (STEMI) probabilities: <0.5; ML (NSTEMI) probabilities: <0.4) and high-risk (TIMI score: >5; ML (STEMI) probabilities: ≥0.5; ML (NSTEMI) probabilities: ≥0.4).

Hence, the ML models demonstrated better predictive accuracy for mortality among high-risk patients compared to the TIMI risk score. Furthermore, ML models demonstrated the greatest improvement in predicting mortality among NSTEMI patients in the context of air pollution exposure.

Table 4.7: Percentage distribution of patient mortality as classified by TIMI Score and ML models across in-hospital selected features and emergency selected features datasets.

Dataset	Predictive Models	High-Risk Threshold	Low-Risk (%)	High-Risk (%)
In-Hospital Selected Features	TIMI (STEMI)	>5	0.85	19.53
	TIMI (NSTEMI)	>5	4.86	9.38
	RF (STEMI)	>0.5	3.03	25.53
	RF (NSTEMI)	>0.4	1.78	18.37
Emergency Selected Features	TIMI (STEMI)	>5	5.14	30.38
	TIMI (NSTEMI)	>5	6.90	6.01
	XGBoost (STEMI)	>0.5	4.19	31.54
	XGBoost (NSTEMI)	>0.4	2.07	16.85

4.2.2.4 Net Reclassification Index (NRI) Analysis

The ML models had significantly better accuracy as assessed by Net Reclassification Index (NRI). NRI for the in-hospital selected features, the net reclassification for STEMI patients using the RF was 8.71%, as shown in Table 4.8, indicating a statistically improvement over the initial TIMI risk score (p<0.001). The NRI for NSTEMI patients, as shown in Table 4.9, shown that the RF model improved net reclassification by 86.94%, substantially outperforming the original TIMI risk score (p<0.001).

Table 4.8: Net Reclassification Improvement (NRI) of the RF Model compared to the TIMI risk score using the in-hospital selected features dataset. The table depicts the comparative performance of the RF model against the TIMI Risk Score for STEMI patients.

In-hospital Selected Features						
		Number of individuals		**Reclassification**		**Net correctly reclassified (%)**
		Random Forest		**Increased risk**	**Decreased risk**	
		Low risk	**High risk**			
Individuals with events (died) (n = 171)						
	TIMI score			34	26	8/171 = 4.68%
	Low risk	37	34			
	High risk	26	74			
Individuals without events (alive) (n = 2334)						
	TIMI score			124	218	94/2334 = 4.03%
	Low risk	1798	124			
	High risk	218	194			
Net Reclassification Index (NRI)	4.68 + 4.03 = 8.71%					
Z, p-value	$Z = \frac{8.71}{\sqrt{\frac{34+26}{171^2}+\frac{124+218}{2334^2}}} = 189.41$					
	189.41, $p < 0.001$					
Conclusion	It was statistically significant. The predictive power of the RF model was improved as compared to the TIMI Risk Scores Model in predicting the mortality rate of ACS STEMI patients in the presence of air pollution, and the proportion of correct classification increased by 8.71%					

Table 4.9: Net Reclassification Improvement (NRI) of the RF Model Compared to the TIMI Risk Score using the In-Hospital Selected Features Dataset. The table depicts the comparative performance of the RF model against the TIMI Risk Score for NSTEMI Patients.

In-hospital Selected Features						
		Number of individuals		**Reclassification**		**Net correctly reclassified (%)**
		Random Forest		**Increased risk**	**Decreased risk**	
		Low risk	**High risk**			
Individuals with events (died) (n = 86)						
	TIMI score			54	2	52/86 = 95.35%
	Low risk	29	54			
	High risk	2	1			
Individuals without events (alive) (n = 1653)						
	TIMI score			162	23	− 139/1653 = − 0.084
	Low risk	1462	162			
	High risk	23	6			
Net Reclassification Index (NRI)	95.35 + (−0.084) = 86.94%					
Z, p-value	$Z = \frac{86.84}{\sqrt{\frac{54+2}{86^2}+\frac{162+23}{1653^2}}} = 994.7$ 994.7, p < 0.001					
Conclusion	It was statistically significant. The predictive power of the RF model was improved as compared to the TIMI Risk Scores Model in predicting the mortality rate of ACS NSTEMI patients in the presence of air pollution, and the proportion of correct classification increased by 86.94%					

While the emergency features, the XGBoost model exhibited the best performance. As shown in Table 4.10, the net reclassification of STEMI patients improved by 5.95%, surpassing the original TIMI risk score. In contrast, the net reclassification improvement for NSTEMI was 50.75% (Table 4.11), a significant improvement compared to the original TIMI risk score ($p<0.001$). Concerning the impact of air pollution, all the ML models outperformed the TIMI risk score.

Table 4.10: Net Reclassification Improvement (NRI) of the XGBoost model compared to the TIMI Risk Score using the emergency selected features dataset. The table depicts the comparative performance of the XGBoost model against the TIMI Risk Score for STEMI patients.

Emergency Selected Features						
		Number of individuals		**Reclassification**		**Net correctly reclassified (%)**
		XGBoost		**Increased risk**	**Decreased risk**	
		Low risk	**High risk**			
Individuals with events (died) (n = 419)						
	TIMI score			70	41	29/419 = 6.92%
	Low risk	74	70			
	High risk	41	234			
Individuals without events (alive) (n = 3287)						
	TIMI score			249	217	− 32/3287 = −0.97%
	Low risk	2410	249			
	High risk	217	411			
Net Reclassification Index (NRI)	6.92 + (−0.97) = 5.95%					
Z, p-value	$Z = \frac{5.95}{\sqrt{\frac{70+41}{419^2} + \frac{249+217}{3287^2}}} = 228.95$ 228.95, p < 0.001					
Conclusion	It was statistically significant. The predictive power of the XGBoost model was improved as compared to the TIMI Risk Scores Model in predicting the mortality rate of ACS STEMI patients in the presence of air pollution, and the proportion of correct classification increased by 5.95%					

Table 4.11: Net Reclassification Improvement (NRI) of the XGBoost Model compared to the TIMI Risk Score using the emergency selected features dataset. The table depicts the comparative performance of the XGBoost model against the TIMI Risk Score for NSTEMI patients.

Emergency Selected Features						
		Number of individuals		**Reclassification**		**Net correctly reclassified (%)**
		XGBoost		**Increased risk**	**Decreased risk**	
		Low risk	**High risk**			
Individuals with events (died) (n = 183)						
	TIMI score			117	1	116/183 = 63.69%
	Low risk	62	117			
	High risk	1	3			
Individuals without events (alive) (n = 2851)						
	TIMI score			408	39	− 369/2851 = −12.94
	Low risk	2389	408			
	High risk	39	15			
Net Reclassification Index (NRI)	63.69 + (−12.94) = 50.75%					
Z, p-value	$Z = \frac{50.75}{\sqrt{\frac{117+1}{183^2}+\frac{408+39}{2851^2}}} = 848.37$ 848.37, p < 0.001					
Conclusion	It was statistically significant. The predictive power of the XGBoost model was improved as compared to the TIMI Risk Scores Model in predicting the mortality rate of ACS NSTEMI patients in the presence of air pollution, and the proportion of correct classification increased by 50.75%					

4.3 Web System Prototype

This section focuses on the web-based system that integrates the optimal ML models outlined in Sections 4.1 and 4.2. For ACS hospitalization prediction, the RF model outperforms other ML algorithms that are utilized in this study, as for predicting ACS mortality rates, XGBoost shows better performance among the evaluated ML algorithms. For the classification models that predict the mortality risk for ACS patients, the RF model has better performance with the in-hospital selected features, whereas XGBoost demonstrated better performance using the emergency selected features dataset.

An overview of the system's functionality, including its outcomes and user interface features, is presented, and discussed in Section 4.3.1. In addition, section 4.3.2 will elaborate on the results of a usability evaluation conducted using the System Usability Scale (SUS).

The developed web system is known as MyHeart Air. It is an AI-powered tool designed to predict cardiovascular outcomes by integrating both cardiac and air quality data. By considering environmental factors, this tool provides a more comprehensive and contextualized prediction tailored to the Malaysian population.

4.3.1 Web System Design and Functionality

Section 4.3.1 presents an overview of the design and functionality of 'MyHeart ACS Air'. The web system has been designed with intuitive navigation to ensure easy use by hospital administrators and healthcare professionals.

Various features have been integrated into the system, including calculators for predicting patient outcomes in the presence of air pollution, such as ACS hospitalization and mortality rates, and ACS patients’ mortality risk in the presence of air pollution. The design of the web-based system includes an interactive element, enabling users to manipulate input data and instantly receive corresponding predicted outcomes. In addition, the system includes a comprehensive patient database. This feature enables users to efficiently manage and update the records of ACS patients.

4.3.1.1 MyHeart ACS Air Homepage

The homepage of ‘MyHeart ACS Air’ web system is shown in Figure 4.18 below, where it is the first page users see when they visit the website and prompts the user for registration and login. It hosts a navigation bar at the top, which contains links to different sections of the system such as “Home”, “Login” and “About”.

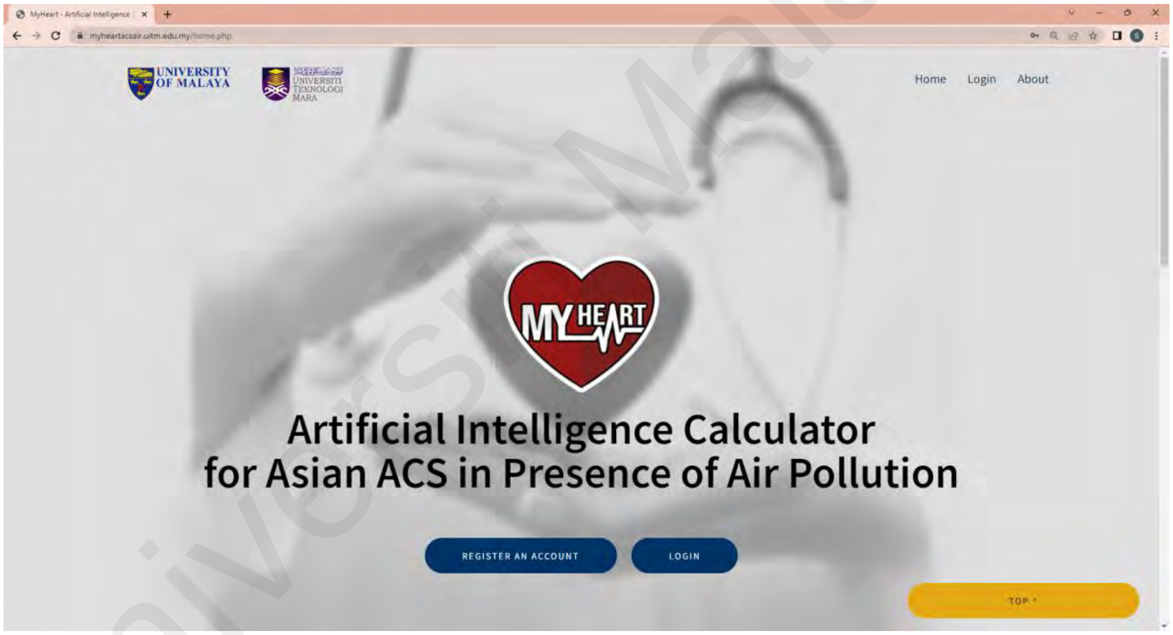

MyHeart - Artificial Intelligence
myheartacsair.uitm.edu.my/home.php
UNIVERSITY OF MALAYA
UNIVERSITI TEKNOLOGI MARA
Home
Login
About
MY HEART
Artificial Intelligence Calculator
for Asian ACS in Presence of Air Pollution
REGISTER AN ACCOUNT
LOGIN
TOP

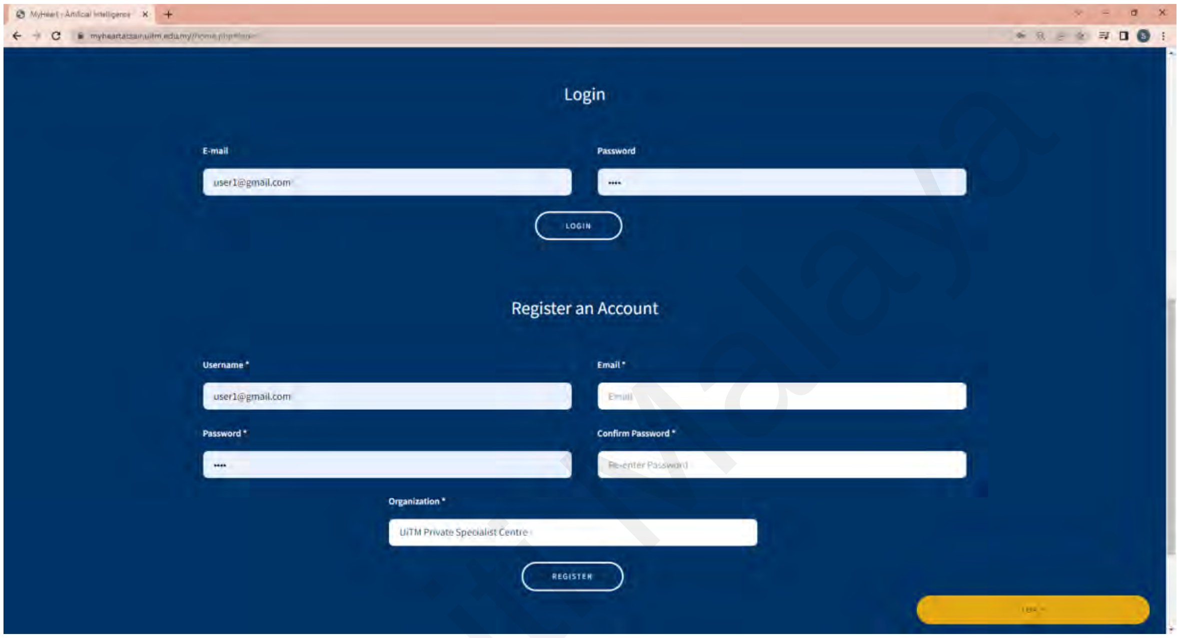
Figure 4.18: MyHeart ACS Air homepage

4.3.1.2 MyHeart ACS Air Registration and Login

The ‘MyHeart ACS Air’ system requires users to register and login to gain full access to its features. As seen in Figure 4.19 shows the login and registration page.

Existing users simply require entering their registered email and password to access the system. The entered details match the records, and the account has been approved by the admin, the users are granted access to the system’s features.

For new users, during the registration process, users are prompted to provide their username, email, password, confirmation of the password, and their respective organizations. Once the registration is completed users have to wait for admin approval of their account. This verification process enhances the system's security and prevents unauthorized access.



Login

E-mail: user1@gmail.com

Password: ••••

LOGIN

Register an Account

Username *: user1@gmail.com

Email *: Email

Password *: ••••

Confirm Password *: Re-enter Password

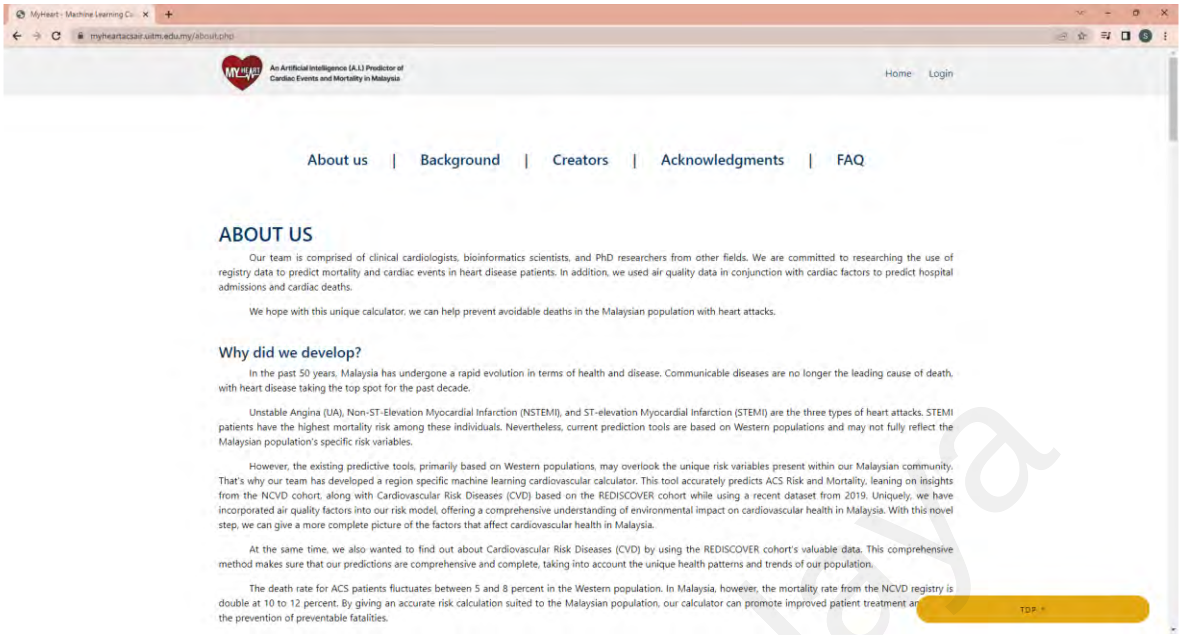
Organization *: UiTM Private Specialist Centre

REGISTER

Figure 4.19: Login and registration of 'MyHeart ACS Air' for accessing the system features.

4.3.1.3 MyHeart ACS Air About Us and FAQ

The 'About Us' page is accessible via the navigation bar of the MyHeart ACS Air, it informs the users about the background of the system, illustrated in Figure 4.20 below. The page outlines the reasons for its development, information about its creators, acknowledgment for the assistance received during the development process and a FAQ section.


MyHeart - Machine Learning Ca
myheartacsair.uitm.edu.my/about.php
MY HEART
An Artificial Intelligence (A.I.) Predictor of Cardiac Events and Mortality in Malaysia
Home
Login
About us | Background | Creators | Acknowledgments | FAQ
TOP ↑

ABOUT US

Our team is comprised of clinical cardiologists, bioinformatics scientists, and PhD researchers from other fields. We are committed to researching the use of registry data to predict mortality and cardiac events in heart disease patients. In addition, we used air quality data in conjunction with cardiac factors to predict hospital admissions and cardiac deaths.

We hope with this unique calculator, we can help prevent avoidable deaths in the Malaysian population with heart attacks.

Why did we develop?

In the past 50 years, Malaysia has undergone a rapid evolution in terms of health and disease. Communicable diseases are no longer the leading cause of death, with heart disease taking the top spot for the past decade.

Unstable Angina (UA), Non-ST-Elevation Myocardial Infarction (NSTEMI), and ST-elevation Myocardial Infarction (STEMI) are the three types of heart attacks. STEMI patients have the highest mortality risk among these individuals. Nevertheless, current prediction tools are based on Western populations and may not fully reflect the Malaysian population's specific risk variables.

However, the existing predictive tools, primarily based on Western populations, may overlook the unique risk variables present within our Malaysian community. That's why our team has developed a region specific machine learning cardiovascular calculator. This tool accurately predicts ACS Risk and Mortality, leaning on insights from the NCVD cohort, along with Cardiovascular Risk Diseases (CVD) based on the REDISCOVER cohort while using a recent dataset from 2019. Uniquely, we have incorporated air quality factors into our risk model, offering a comprehensive understanding of environmental impact on cardiovascular health in Malaysia. With this novel step, we can give a more complete picture of the factors that affect cardiovascular health in Malaysia.

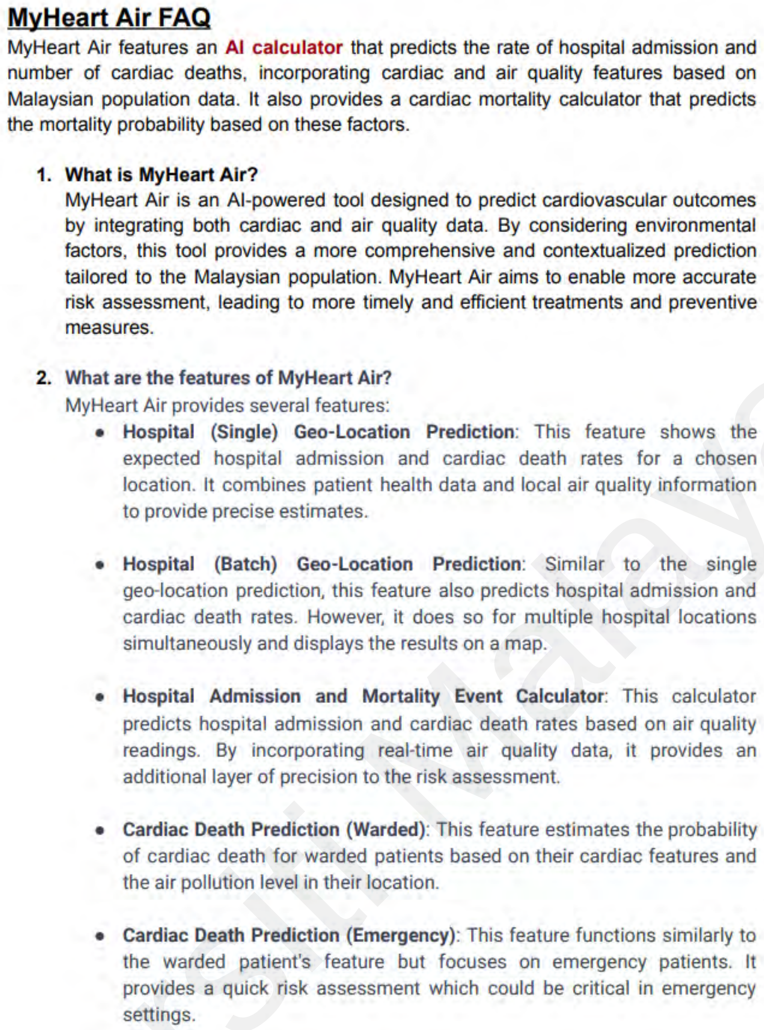
At the same time, we also wanted to find out about Cardiovascular Risk Diseases (CVD) by using the REDISCOVER cohort's valuable data. This comprehensive method makes sure that our predictions are comprehensive and complete, taking into account the unique health patterns and trends of our population.

The death rate for ACS patients fluctuates between 5 and 8 percent in the Western population. In Malaysia, however, the mortality rate from the NCVD registry is double at 10 to 12 percent. By giving an accurate risk calculation suited to the Malaysian population, our calculator can promote improved patient treatment an

the prevention of preventable fatalities.

Figure 4.20: About Us page of 'MyHeart ACS Air' provide the reasons about the system's development, information of the developers, acknowledgements, and FAQ.

The FAQ section is included in the 'About Us' page. The FAQ covers the information and reliability about the 'MyHeart ACS Air' system's features, the accuracy, and the Receiver Operating Characteristic (ROC) incorporated into the system. The detailed version of the FAQ is included in the Appendix G of the thesis. Figure 4.21 presents a snapshot of the FAQ available within the 'About Us' page.



MyHeart Air FAQ

MyHeart Air features an **AI calculator** that predicts the rate of hospital admission and number of cardiac deaths, incorporating cardiac and air quality features based on Malaysian population data. It also provides a cardiac mortality calculator that predicts the mortality probability based on these factors.

1. **What is MyHeart Air?**
   MyHeart Air is an AI-powered tool designed to predict cardiovascular outcomes by integrating both cardiac and air quality data. By considering environmental factors, this tool provides a more comprehensive and contextualized prediction tailored to the Malaysian population. MyHeart Air aims to enable more accurate risk assessment, leading to more timely and efficient treatments and preventive measures.

2. **What are the features of MyHeart Air?**
   MyHeart Air provides several features:
   - **Hospital (Single) Geo-Location Prediction**: This feature shows the expected hospital admission and cardiac death rates for a chosen location. It combines patient health data and local air quality information to provide precise estimates.
   - **Hospital (Batch) Geo-Location Prediction**: Similar to the single geo-location prediction, this feature also predicts hospital admission and cardiac death rates. However, it does so for multiple hospital locations simultaneously and displays the results on a map.
   - **Hospital Admission and Mortality Event Calculator**: This calculator predicts hospital admission and cardiac death rates based on air quality readings. By incorporating real-time air quality data, it provides an additional layer of precision to the risk assessment.
   - **Cardiac Death Prediction (Warded)**: This feature estimates the probability of cardiac death for warded patients based on their cardiac features and the air pollution level in their location.
   - **Cardiac Death Prediction (Emergency)**: This feature functions similarly to the warded patient's feature but focuses on emergency patients. It provides a quick risk assessment which could be critical in emergency settings.

Figure 4.21: A snapshot of the FAQ for 'MyHeart ACS Air', addressing the frequently asked questions about the system's features, the model accuracy, and the ROC.

4.3.1.4 MyHeart ACS Air Dashboard

The 'MyHeart ACS Air' dashboard works as the system's primary hub, providing hospital administrators and healthcare professionals with various AI-powered predictive functionalities (Figure 4.22). The features of the system are as follows:

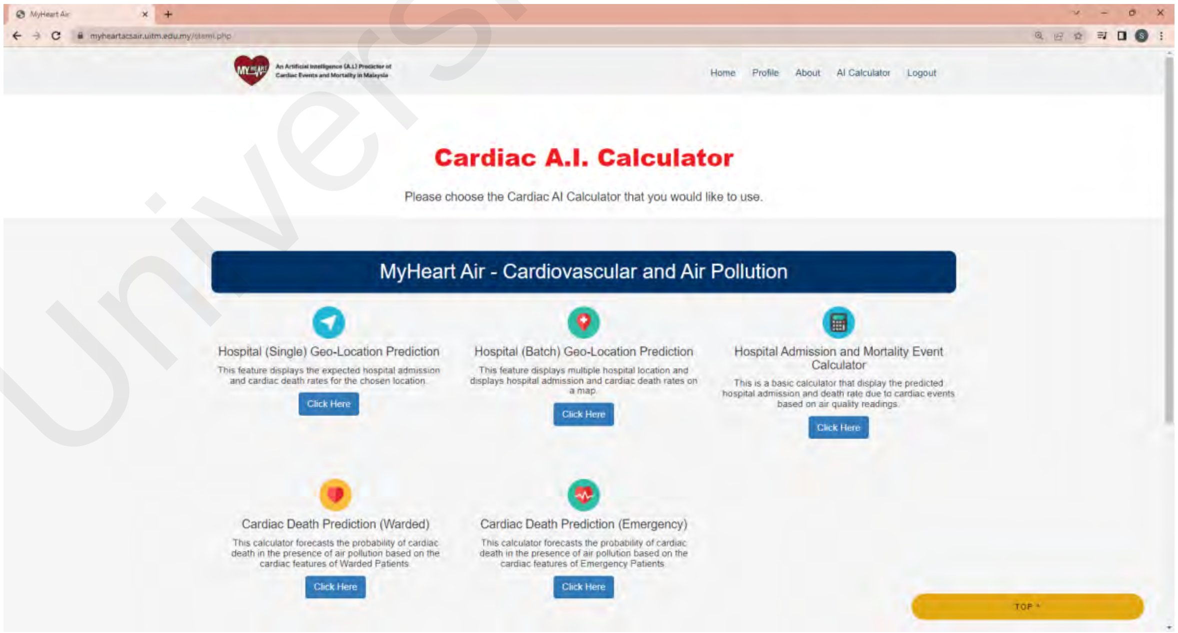
1. Hospital (Single) Geo-Location Prediction: This feature combines patient health data and local air quality information to estimate expected ACS hospitalization and mortality rates for a single selected location.

2. Hospital (Multiple) Geo-Location Prediction: Working similarly to the single geo-location prediction, this feature provides estimates for multiple hospital locations at once, further displaying the results on an interactive map for easy visualization and comparison.

3. ACS hospitalization and Mortality Event Calculator: This calculator presents an enhanced risk assessment by predicting ACS hospitalization and mortality rates based on air quality readings.

4. ACS Mortality Prediction (In-hospital): This feature aims at aiding in ACS patient care by estimating the probability of cardiac mortality for in-patients, based on their individual cardiac features and the level of air pollution in their location.

5. ACS Mortality Prediction (Emergency): Tailored specifically for emergency settings, this feature gives a quick risk assessment of cardiac mortality for emergency ACS patients, offering potentially lifesaving insights to healthcare professionals.


MyHeart Air
myheartacsair.uitm.edu.my/stemi.php
An Artificial Intelligence (A.I.) Predictor of Cardiac Events and Mortality in Malaysia
Home
Profile
About
AI Calculator
Logout
Click Here
Click Here
Click Here
Click Here
Click Here
TOP ^

Cardiac A.I. Calculator

Please choose the Cardiac AI Calculator that you would like to use.

MyHeart Air - Cardiovascular and Air Pollution

Hospital (Single) Geo-Location Prediction

This feature displays the expected hospital admission and cardiac death rates for the chosen location.

Hospital (Batch) Geo-Location Prediction

This feature displays multiple hospital location and displays hospital admission and cardiac death rates on a map.

Hospital Admission and Mortality Event Calculator

This is a basic calculator that display the predicted hospital admission and death rate due to cardiac events based on air quality readings.

Cardiac Death Prediction (Warded)

This calculator forecasts the probability of cardiac death in the presence of air pollution based on the cardiac features of Warded Patients.

Cardiac Death Prediction (Emergency)

This calculator forecasts the probability of cardiac death in the presence of air pollution based on the cardiac features of Emergency Patients.

Figure 4.22: Overview of the 'MyHeart ACS Air' interface, presenting various features based on artificial intelligence predictive functionalities available for user interaction.

4.3.1.5 MyHeart ACS Air Calculators

The "MyHeart ACS Air" system features several custom tools, each of which is created to satisfy the requirements of a certain user. These calculators serve as the system's foundation, providing users with personalized predictions using ML. Their primary goal is to predict ACS hospitalization rates, mortality rates, and the mortality risk for ACS patients in the presence of air pollution.

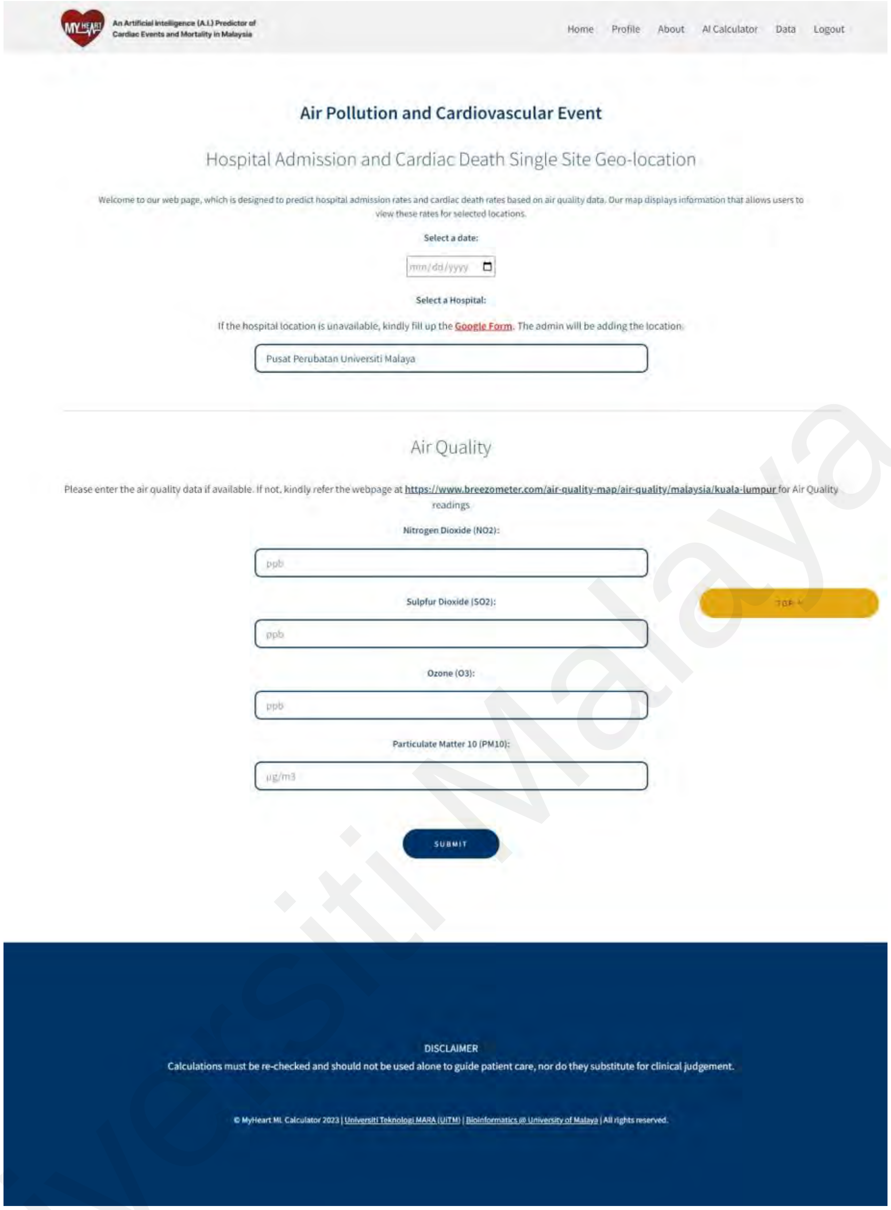
While each calculator has unique functions, they all contribute to a comprehensive system that attempts to support medical practitioners in providing better patient care. The subsections that follow will provide a full description of each calculator's particular functioning.

4.3.1.5.1 Hospital (Single) Geo-Location Prediction

The 'Hospital (Single) Geo-place Prediction' is a feature that enables users to predict ACS hospitalization rates and mortality rates based on air quality data for a chosen location. The only input required for this feature's user interface is the date and hospital location, which can be chosen from a predetermined list in the database.

User-level access to this feature allows for data input via Google Forms, while admin-level access provides additional control, including the ability to add new locations to the database. This design decision maintains the integrity of the location data by minimizing potential user-induced discrepancies.

Once the location is selected, users are then asked to input air quality data, including measures of nitrogen oxides, sulfur dioxide, ozone, and particulate matter 10. For users unsure of the air quality readings, a helpful link to an external webpage (https://www.breezometer.com/air-quality-map/air-quality/malaysia/kuala-lumpur) is provided for reference. Figure 4.23 below shows the input page interface.



An Artificial Intelligence (A.I.) Predictor of Cardiac Events and Mortality in Malaysia
Home
Profile
About
AI Calculator
Data
Logout
TOP
SUBMIT

Air Pollution and Cardiovascular Event

Hospital Admission and Cardiac Death Single Site Geo-location

Welcome to our web page, which is designed to predict hospital admission rates and cardiac death rates based on air quality data. Our map displays information that allows users to view these rates for selected locations.

Select a date: mm/dd/yyyy

Select a Hospital:

If the hospital location is unavailable, kindly fill up the Google Form. The admin will be adding the location.

Pusat Perubatan Universiti Malaya

Air Quality

Please enter the air quality data if available. If not, kindly refer the webpage at https://www.breezometer.com/air-quality-map/air-quality/malaysia/kuala-lumpur for Air Quality readings

Nitrogen Dioxide (NO2): ppb

Sulpfur Dioxide (SO2): ppb

Ozone (O3): ppb

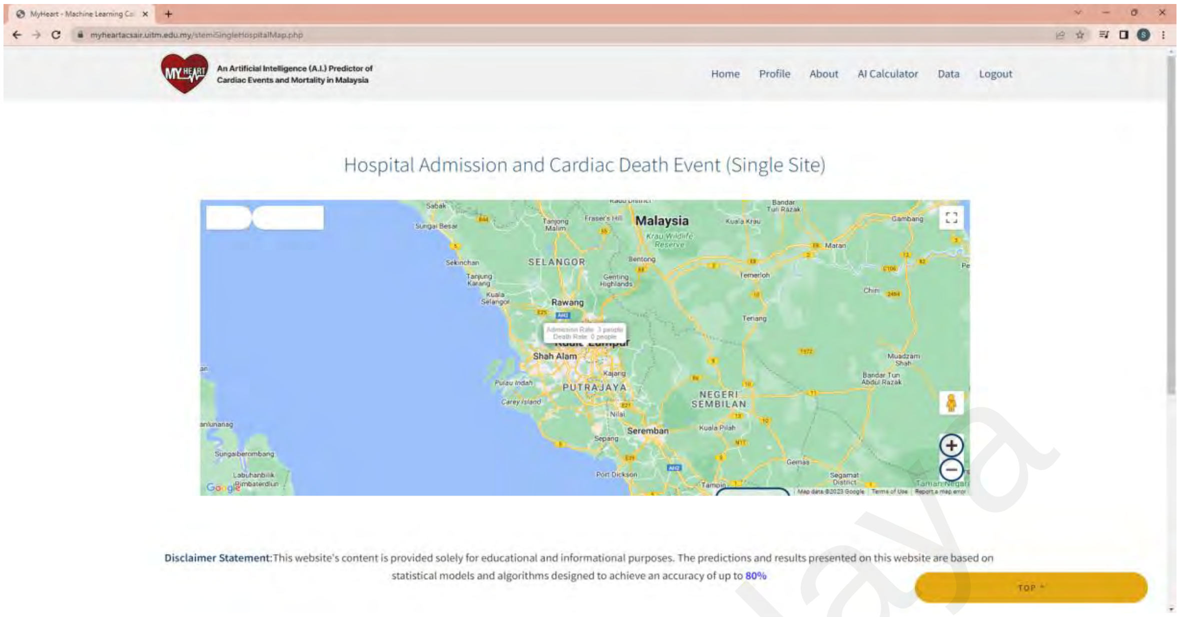
Particulate Matter 10 (PM10): µg/m3

DISCLAIMER
Calculations must be re-checked and should not be used alone to guide patient care, nor do they substitute for clinical judgement.

© MyHeart ML Calculator 2023 | Universiti Teknologi MARA (UiTM) | Bioinformatics @ University of Malaya | All rights reserved.

Figure 4.23: The Hospital (Single) Geo-Location Prediction interface prompting users to input date, location, and air quality data.

Once the data is entered, the system generates a figure that displays the selected location, as well as the predicted hospitalization rate and mortality rate for ACS patients in the presence of the entered air quality data as presented in Figure 4.24.


MyHeart - Machine Learning Ca
myheartacsair.uitm.edu.my/stemiSingleHospitalMap.php
An Artificial Intelligence (A.I.) Predictor of Cardiac Events and Mortality in Malaysia
Home
Profile
About
AI Calculator
Data
Logout
Malaysia
SELANGOR
Rawang
Kuala Lumpur
Shah Alam
PUTRAJAYA
NEGERI SEMBILAN
Seremban
TOP

Hospital Admission and Cardiac Death Event (Single Site)

Admission Rate 3 people
Death Rate 0 people

Disclaimer Statement: This website's content is provided solely for educational and informational purposes. The predictions and results presented on this website are based on statistical models and algorithms designed to achieve an accuracy of up to 80%

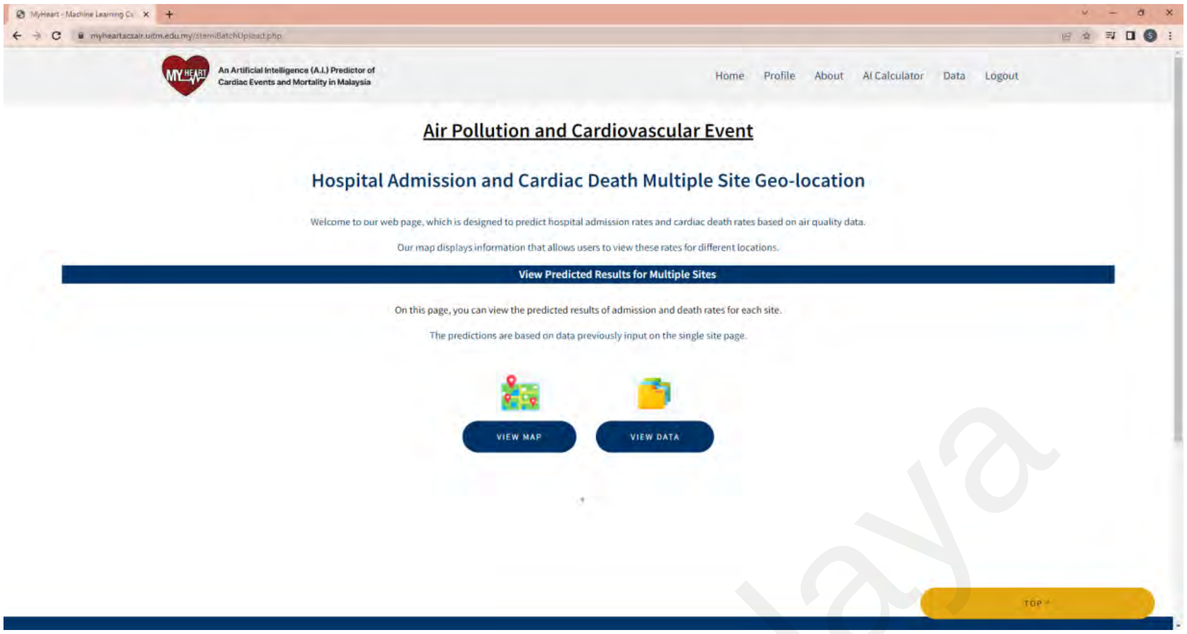
Figure 4.24: The Hospital (Single) Geo-Location Prediction calculator output, showing the chosen location and the predicted hospitalization and mortality rates for ACS patients.

4.3.1.5.2 Hospital (Multiple) Geo-Location Prediction

The 'Hospital (Multiple) Geo-Location Prediction' is designed to provide users with visual representations of ACS hospitalization rates and mortality rates for multiple hospital locations simultaneously, the user interface is shown in Figure 4.25.

The Hospital (Multiple) Geo-Location Prediction page provides users with two main functionalities, which are 'VIEW MAP' and 'VIEW DATA'. The 'VIEW MAP' option enables users to view the geographical distribution of predicted ACS hospitalization and mortality rates for different hospital sites. The predictions displayed on the map are based on the data previously input on the single hospital location prediction page, as shown in Figure 4.26.

On the other hand, the 'VIEW DATA' option allows users to inspect the data that has been input and stored in the database. This includes information pertaining to the various hospital locations and corresponding air quality data. Further details regarding the management and accessibility of this stored data are discussed in Section 4.3.1.6.2 'Hospital Location and Air Quality Database'.



MyHeart - Machine Learning Ca
myheartacsair.uitm.edu.my//stemiBatchUpload.php
MY HEART
Home
Profile
About
AI Calculator
Data
Logout
VIEW MAP
VIEW DATA
TOP

An Artificial Intelligence (A.I.) Predictor of Cardiac Events and Mortality in Malaysia

Air Pollution and Cardiovascular Event

Hospital Admission and Cardiac Death Multiple Site Geo-location

Welcome to our web page, which is designed to predict hospital admission rates and cardiac death rates based on air quality data.

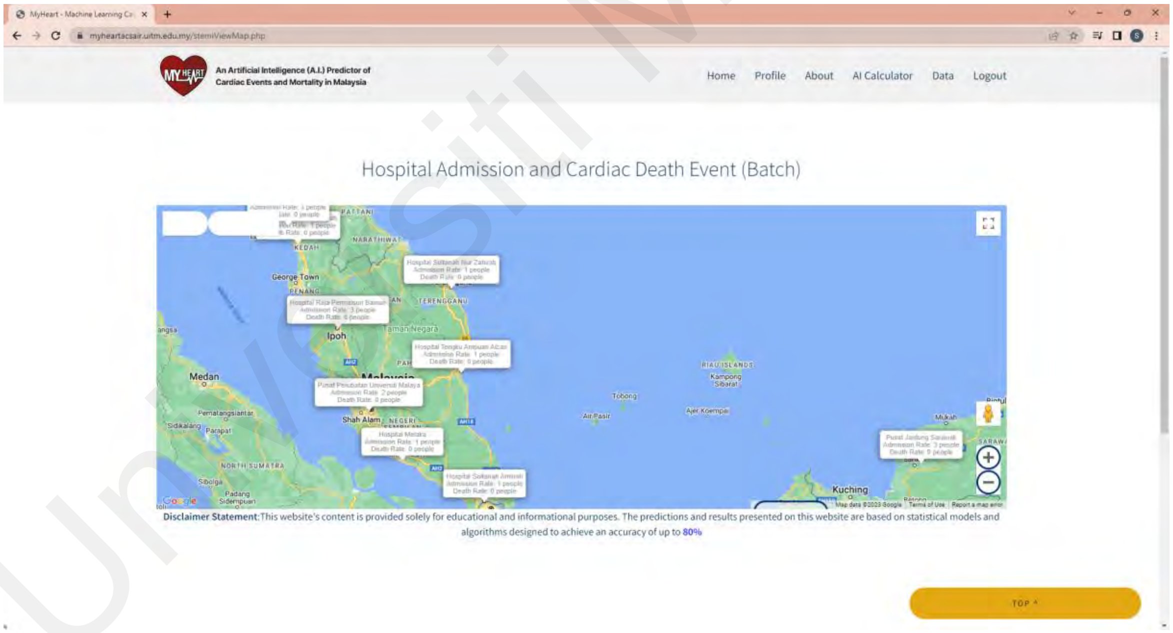
Our map displays information that allows users to view these rates for different locations.

View Predicted Results for Multiple Sites

On this page, you can view the predicted results of admission and death rates for each site.

The predictions are based on data previously input on the single site page.

Figure 4.25: User interface of the 'Hospital (Multiple) Geo-Location Prediction' in 'MyHeart ACS Air' System, providing users with options to 'VIEW MAP' or 'VIEW DATA'.



An Artificial Intelligence (A.I.) Predictor of Cardiac Events and Mortality in Malaysia

Hospital Admission and Cardiac Death Event (Batch)

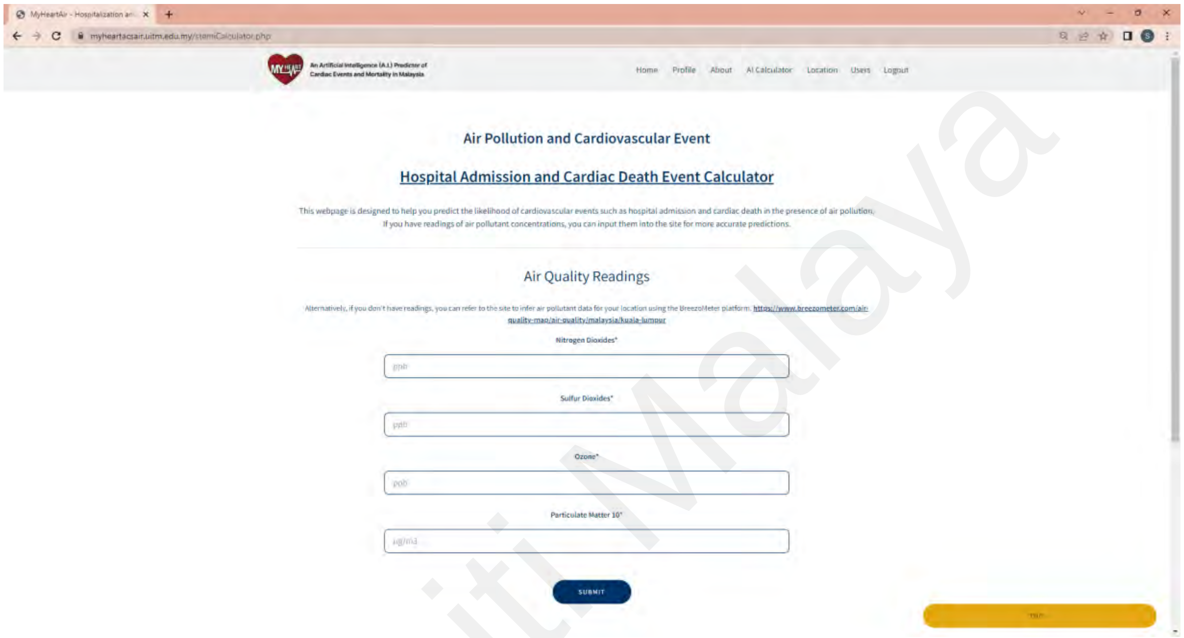
Disclaimer Statement:This website's content is provided solely for educational and informational purposes. The predictions and results presented on this website are based on statistical models and algorithms designed to achieve an accuracy of up to 80%

Figure 4.26: Predicted ACS hospitalization and mortality rates for multiple hospital locations displayed on the Google Map.

4.3.1.5.3 Cardiac Hospitalization and Mortality Event Calculator

The 'ACS hospitalization and Mortality Event Calculator' is a basic calculator that displays the predicted cardiac hospitalization and mortality rate due to ACS events based on air quality readings. This tool is specifically calibrated to utilize air quality data from

Malaysia. Therefore, the generated predictions are particularly relevant and applicable to healthcare scenarios in the Malaysian context. Figure 4.27 and figure 4.28 below depict the user interface, which prompts users to input air quality data, and the subsequent predicted outcomes generated by the ML algorithms.



MyHeartAir - Hospitalization an
myheartacsair.uitm.edu.my/stemiCalculator.php
An Artificial Intelligence (A.I.) Predictor of Cardiac Events and Mortality in Malaysia
Home
Profile
About
AI Calculator
Location
Users
Logout

Air Pollution and Cardiovascular Event

Hospital Admission and Cardiac Death Event Calculator

This webpage is designed to help you predict the likelihood of cardiovascular events such as hospital admission and cardiac death in the presence of air pollution. If you have readings of air pollutant concentrations, you can input them into the site for more accurate predictions.

Air Quality Readings

Alternatively, if you don't have readings, you can refer to the site to infer air pollutant data for your location using the BreezoMeter platform. https://www.breezometer.com/air-quality-map/air-quality/malaysia/kuala-lumpur

Nitrogen Dioxides*

ppb

Sulfur Dioxides*

ppb

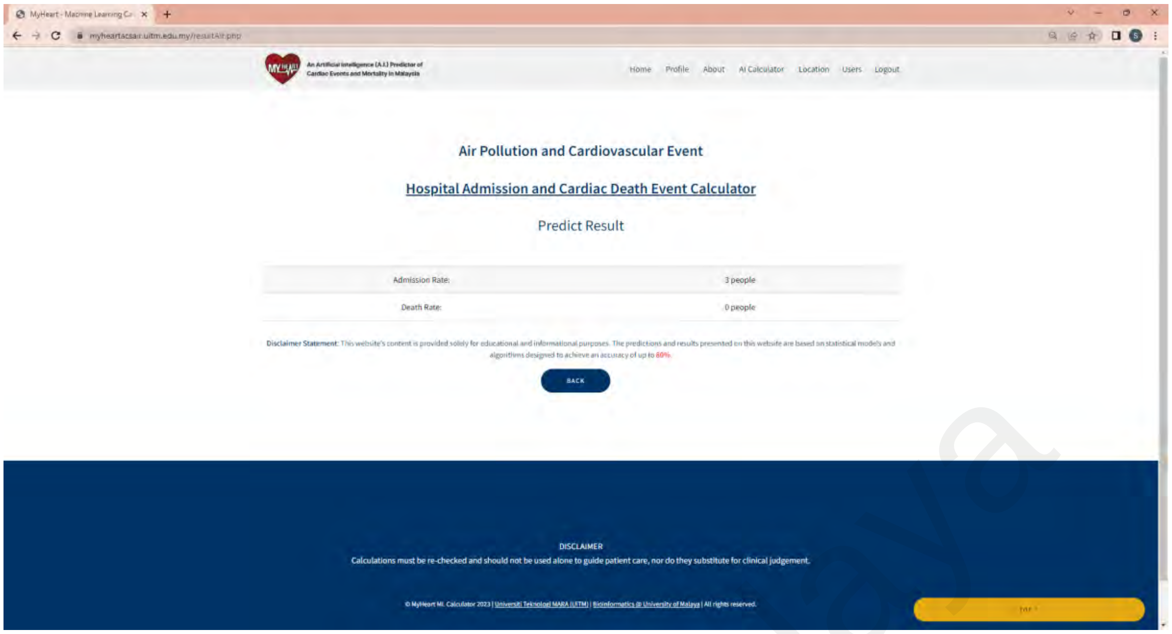
Ozone*

ppb

Particulate Matter 10*

ug/m3

Figure 4.27: User interface of the 'ACS Hospitalization and Mortality Event Calculator' allows users to input air quality readings.


MyHeart - Machine Learning Ca
myheartacsair.uitm.edu.my/resultAir.php
An Artificial Intelligence (A.I.) Predictor of Cardiac Events and Mortality in Malaysia
Home
Profile
About
AI Calculator
Location
Users
Logout
BACK
TOP

Air Pollution and Cardiovascular Event

Hospital Admission and Cardiac Death Event Calculator

Predict Result

Admission Rate:	3 people
Death Rate:	0 people

Disclaimer Statement: This website's content is provided solely for educational and informational purposes. The predictions and results presented on this website are based on statistical models and algorithms designed to achieve an accuracy of up to 60%.

DISCLAIMER

Calculations must be re-checked and should not be used alone to guide patient care, nor do they substitute for clinical judgement.

© MyHeart ML Calculator 2023 | Universiti Teknologi MARA (UiTM) | Bioinformatics @ University of Malaya | All rights reserved.

Figure 4.28: The result is display of predicted ACS hospitalization and ACS mortality rates, generated by 'ACS Hospitalization and Mortality Event Calculator' after user provide inputs air quality readings.

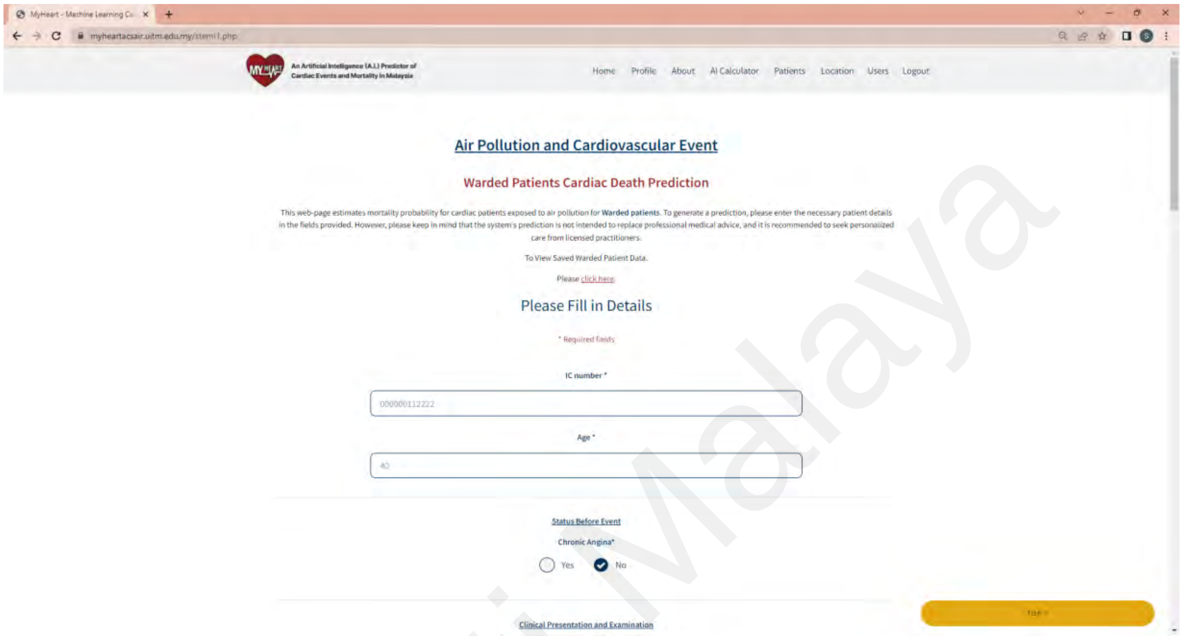
4.3.1.5.4 ACS Mortality Prediction (In-Hospital)

The 'ACS Mortality Prediction (In-Hospital)' function is a specialized tool within the 'MyHeart ACS Air' web system designed specifically for hospitalized ACS patients that estimates the probability of mortality among ACS patients admitted to the hospital, particularly in the context of air pollution.

This feature provides a user interface that prompts medical personnel to input ACS patient details including patient details, status before event, clinical presentation and examination, baseline investigation, electrocardiography, invasive therapeutic procedure, pharmacological therapy, and air quality readings. The snapshot of the user interface is shown in Figure 4.29.

The web system with integrated ML model processes the submitted data to estimate the risk of mortality of ACS patients in the presence of the specified air pollution levels. The expected output is displayed as a percentage, the risk percentage above 50% is considered as high-risk patient, as shown in Figure 4.30 below.

The high-risk indicator alerts medical personnel to be aware of the patients. By providing an accurate assessment of the potential risk incurred by a patient, it is possible to make better care decisions, eventually enhancing patient safety and health outcomes.



MyHeart - Machine Learning Ca
myheartacsair.uitm.edu.my/stemi1.php
An Artificial Intelligence (A.I.) Predictor of Cardiac Events and Mortality in Malaysia
Home Profile About AI Calculator Patients Location Users Logout
TOP

Air Pollution and Cardiovascular Event

Warded Patients Cardiac Death Prediction

This web-page estimates mortality probability for cardiac patients exposed to air pollution for **Warded patients**. To generate a prediction, please enter the necessary patient details in the fields provided. However, please keep in mind that the system's prediction is not intended to replace professional medical advice, and it is recommended to seek personalized care from licensed practitioners.

To View Saved Warded Patient Data.

Please click here.

Please Fill in Details

* Required fields

IC number *

000000112222

Age *

40

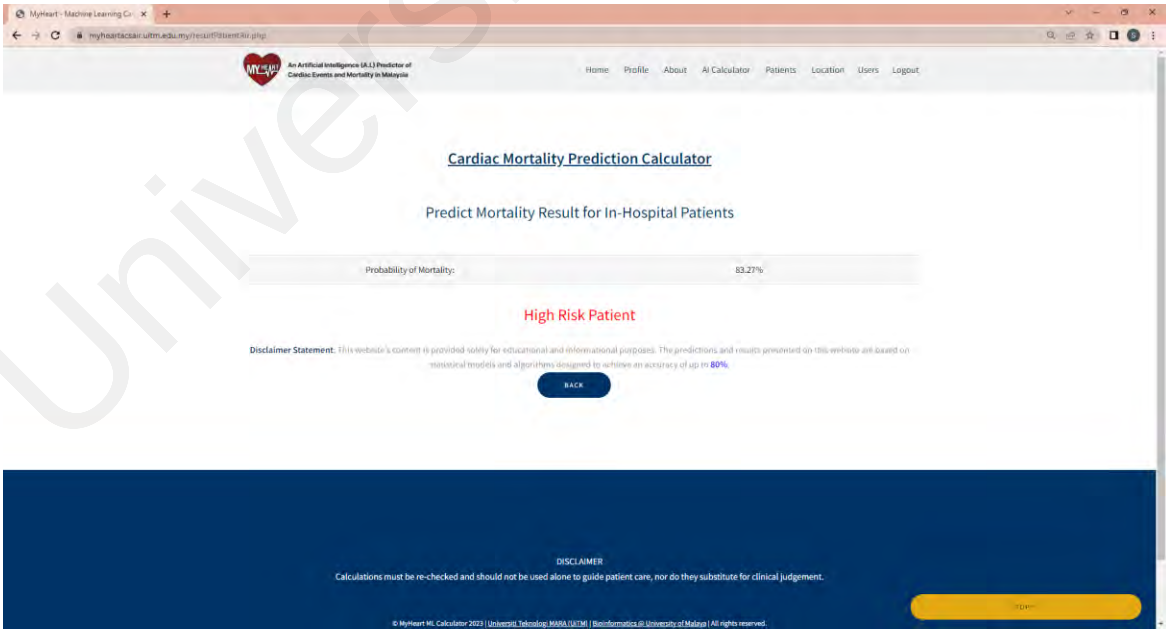
Status Before Event

Chronic Angina*

○ Yes ● No

Clinical Presentation and Examination

Figure 4.29: User interface for input patient data into the 'ACS Mortality Prediction (in-hospital)' page.



MyHeart - Machine Learning Ca
myheartacsair.uitm.edu.my/resultPatientAir.php
An Artificial Intelligence (A.I.) Predictor of Cardiac Events and Mortality in Malaysia
Home Profile About AI Calculator Patients Location Users Logout
BACK
TOP

Cardiac Mortality Prediction Calculator

Predict Mortality Result for In-Hospital Patients

Probability of Mortality:	83.27%

High Risk Patient

Disclaimer Statement: This website's content is provided solely for educational and informational purposes. The predictions and results presented on this website are based on statistical models and algorithms designed to achieve an accuracy of up to 80%.

DISCLAIMER

Calculations must be re-checked and should not be used alone to guide patient care, nor do they substitute for clinical judgement.

© MyHeart ML Calculator 2023 | Universiti Teknologi MARA (UiTM) | Bioinformatics @ University of Malaya | All rights reserved.

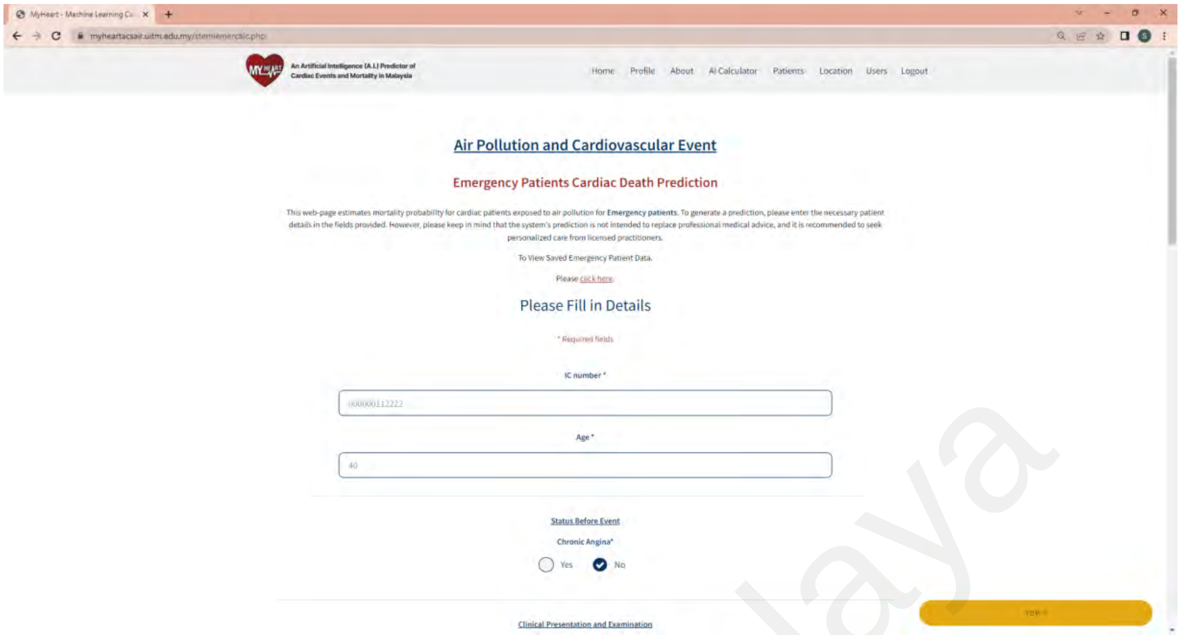
Figure 4.30: "High-Risk Patient" is displayed in the predicted risk of mortality for hospitalized ACS patient in the presence of air pollution, generated by the 'ACS Mortality Prediction (In-hospital)' feature upon user inputs.

4.3.1.5.5 Cardiac Mortality Prediction (Emergency)

The 'ACS Mortality Prediction (Emergency)' tool is another part of the 'MyHeart ACS Air' web system, designed for assessing ACS patients in emergency situations. It is specially designed for emergency hospitalized patients, which requires fewer inputs than the version designed for in-hospital patients. Hence, the users are required to important patient information, such as patient details, status before event, clinical presentation and examination, electrocardiography, pharmacological therapy, and air quality, which excluded the baseline investigation and invasive therapeutic procedure.

Once the data is submitted, the system's integrated ML model processes the information to predict the patient's risk of ACS mortality in the current air pollution conditions (Figure 4.31). The result of the risk prediction is displayed in percentages, similar to the ACS mortality risk calculator for in-hospital patients (Figure 4.32).

This immediate risk indicator assists healthcare professionals in quickly understanding the patient's condition, guiding urgent medical decisions and patient management strategies. This quick assessment tool aids in optimizing patient care in emergency situations.


MyHeart - Machine Learning Ca
myheartacsair.uitm.edu.my/stemiemercalc.php
An Artificial Intelligence (A.I.) Predictor of Cardiac Events and Mortality in Malaysia
Home
Profile
About
AI Calculator
Patients
Location
Users
Logout
TOP

Air Pollution and Cardiovascular Event

Emergency Patients Cardiac Death Prediction

This web-page estimates mortality probability for cardiac patients exposed to air pollution for **Emergency patients**. To generate a prediction, please enter the necessary patient details in the fields provided. However, please keep in mind that the system's prediction is not intended to replace professional medical advice, and it is recommended to seek personalized care from licensed practitioners.

To View Saved Emergency Patient Data.

Please click here.

Please Fill in Details

* Required fields

IC number *

000000112222

Age *

40

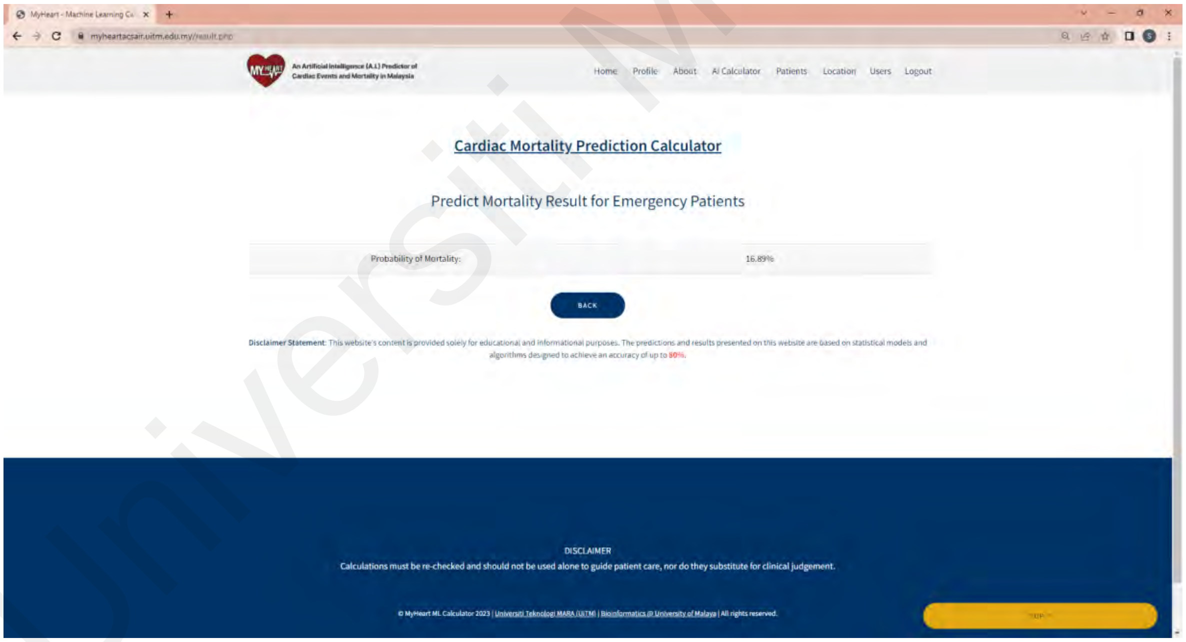
Status Before Event

Chronic Angina*

○ Yes ● No

Clinical Presentation and Examination

Figure 4.31: User interface for input patient data into the 'ACS Mortality Prediction (Emergency)' page.


MyHeart - Machine Learning Ca
myheartacsair.uitm.edu.my/result.php
An Artificial Intelligence (A.I.) Predictor of Cardiac Events and Mortality in Malaysia
Home
Profile
About
AI Calculator
Patients
Location
Users
Logout
BACK
TOP

Cardiac Mortality Prediction Calculator

Predict Mortality Result for Emergency Patients

Probability of Mortality:	16.89%

Disclaimer Statement: This website's content is provided solely for educational and informational purposes. The predictions and results presented on this website are based on statistical models and algorithms designed to achieve an accuracy of up to 80%.

DISCLAIMER

Calculations must be re-checked and should not be used alone to guide patient care, nor do they substitute for clinical judgement.

© MyHeart ML Calculator 2023 | Universiti Teknologi MARA (UiTM) | Bioinformatics @ University of Malaya | All rights reserved.

Figure 4.32: The result page of the 'ACS Mortality Prediction (Emergency)' feature, displaying the predicted mortality risk.

4.3.1.6 MyHeart ACS Air Databases

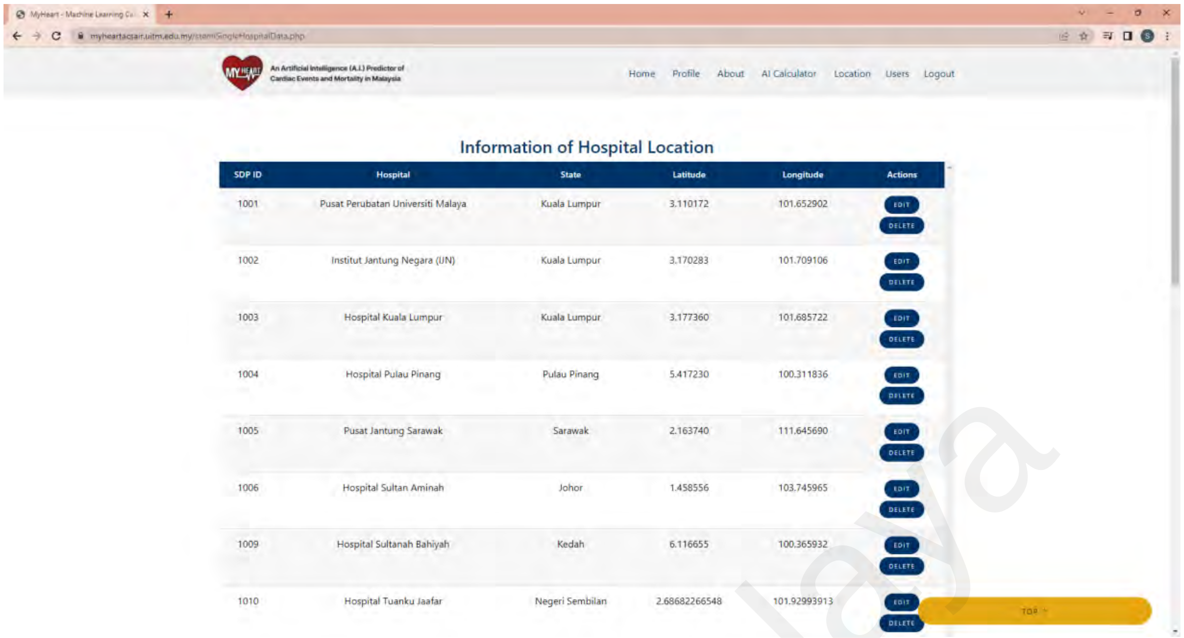
Apart from the calculators that generate predicted results supported by ML, the 'MyHeart ACS Air' web system is supported by data management system, which comprises five primary databases. Each of these databases serves distinct yet interrelated functions, to support the various predictive features of the system.

Each database serves its unique purpose and functionality, to ensure the ‘MyHeart ACS Air’ system provides precise predictions to assist in better healthcare decision making. A detailed description of each database and its functionalities is provided in the subsequent section.

4.3.1.6.1 Hospital Location Database

The hospital location database features geographical information of the various hospital locations. It is primarily managed by administrative-level users to maintain data integrity and ensure accurate location-based predictions.

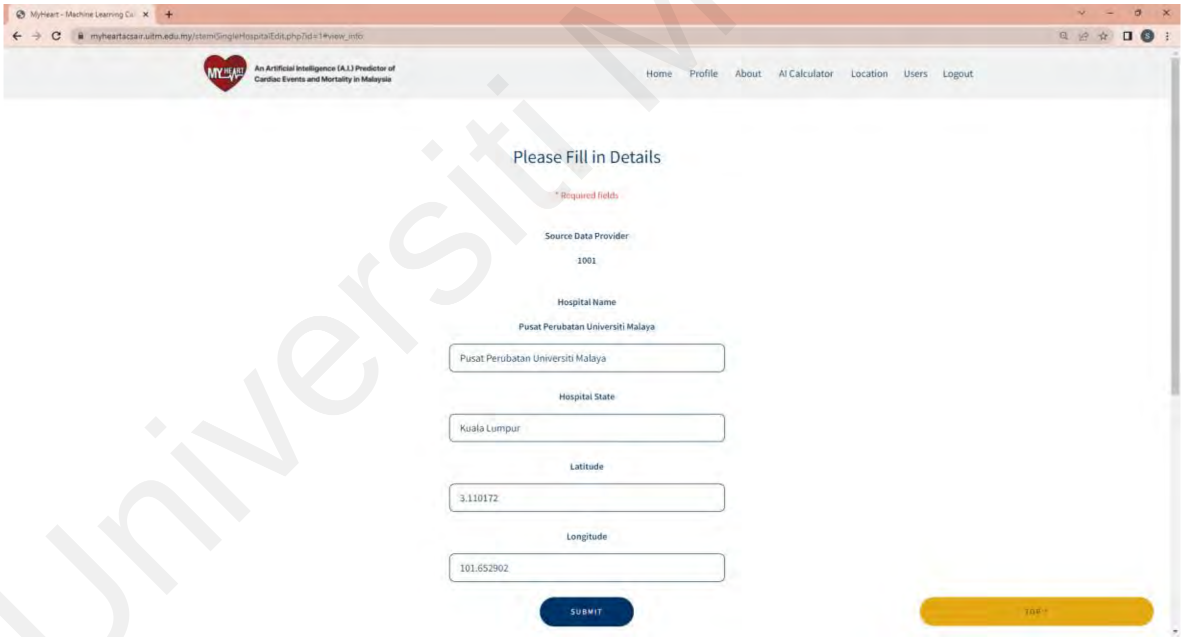
The database contains information related to the Source Data Provider (SDP) ID acquired based on the NCVD-ACS Annual Report, name of the hospital, the state in which it is located and its geographical coordinates (latitude and longitude). Additionally, the interactive database allows the administrator to add, edit, and delete records when necessary. This is to ensure that the geographical location and the system’s predictions are based on the most accurate and updated location data. Figure 4.33 depicts the snapshot of the ‘Hospital Location Database’ page and Figure 4.34 is the ‘Edit Location Data’ page, where user can edit and update the specific location data.


MyHeart - Machine Learning Ca
myheartacsair.uitm.edu.my/stemiSingleHospitalData.php
An Artificial Intelligence (A.I.) Predictor of Cardiac Events and Mortality in Malaysia
Home
Profile
About
AI Calculator
Location
Users
Logout
TOP

Information of Hospital Location

SDP ID	Hospital	State	Latitude	Longitude	Actions
1001	Pusat Perubatan Universiti Malaya	Kuala Lumpur	3.110172	101.652902	EDIT DELETE
1002	Institut Jantung Negara (IJN)	Kuala Lumpur	3.170283	101.709106	EDIT DELETE
1003	Hospital Kuala Lumpur	Kuala Lumpur	3.177360	101.685722	EDIT DELETE
1004	Hospital Pulau Pinang	Pulau Pinang	5.417230	100.311836	EDIT DELETE
1005	Pusat Jantung Sarawak	Sarawak	2.163740	111.645690	EDIT DELETE
1006	Hospital Sultan Aminah	Johor	1.458556	103.745965	EDIT DELETE
1009	Hospital Sultanah Bahiyah	Kedah	6.116655	100.365932	EDIT DELETE
1010	Hospital Tuanku Jaafar	Negeri Sembilan	2.68682266548	101.92993913	EDIT DELETE

Figure 4.33: The 'Hospital Location Database' page, displaying key information related to various hospital locations.


MyHeart - Machine Learning Ca
myheartacsair.uitm.edu.my/stemiSingleHospitalEdit.php?id=1#view_info
An Artificial Intelligence (A.I.) Predictor of Cardiac Events and Mortality in Malaysia
Home
Profile
About
AI Calculator
Location
Users
Logout
SUBMIT
TOP

Please Fill in Details

* Required fields

Source Data Provider

1001

Hospital Name

Pusat Perubatan Universiti Malaya

Pusat Perubatan Universiti Malaya

Hospital State

Kuala Lumpur

Latitude

3.110172

Longitude

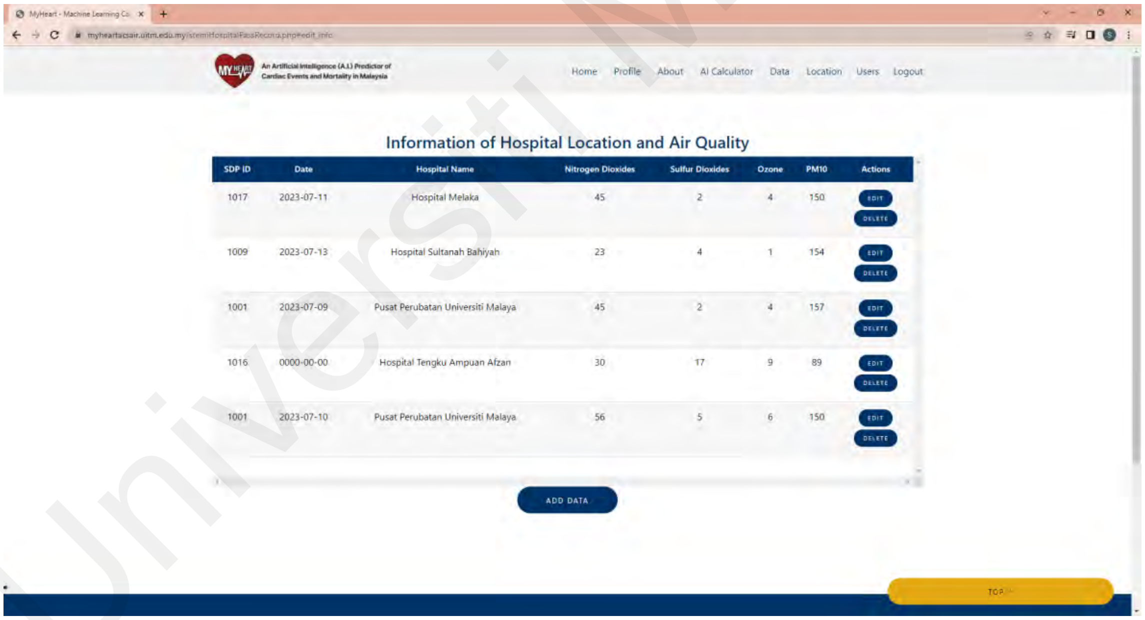
101.652902

Figure 4.34: The 'Edit Location Data' page, accessible only by the admin, allows for vigilant modifications of location data in the database.

4.3.1.6.2 Hospital Location and Air Quality Database

This database combines specific location data with corresponding air quality readings. The ‘Hospital Location and Air Quality Database’ contains key information including the SDP ID, date, hospital name, and air quality readings, such as nitrogen oxides, sulfur dioxide, ozone, and particulate matter 10.

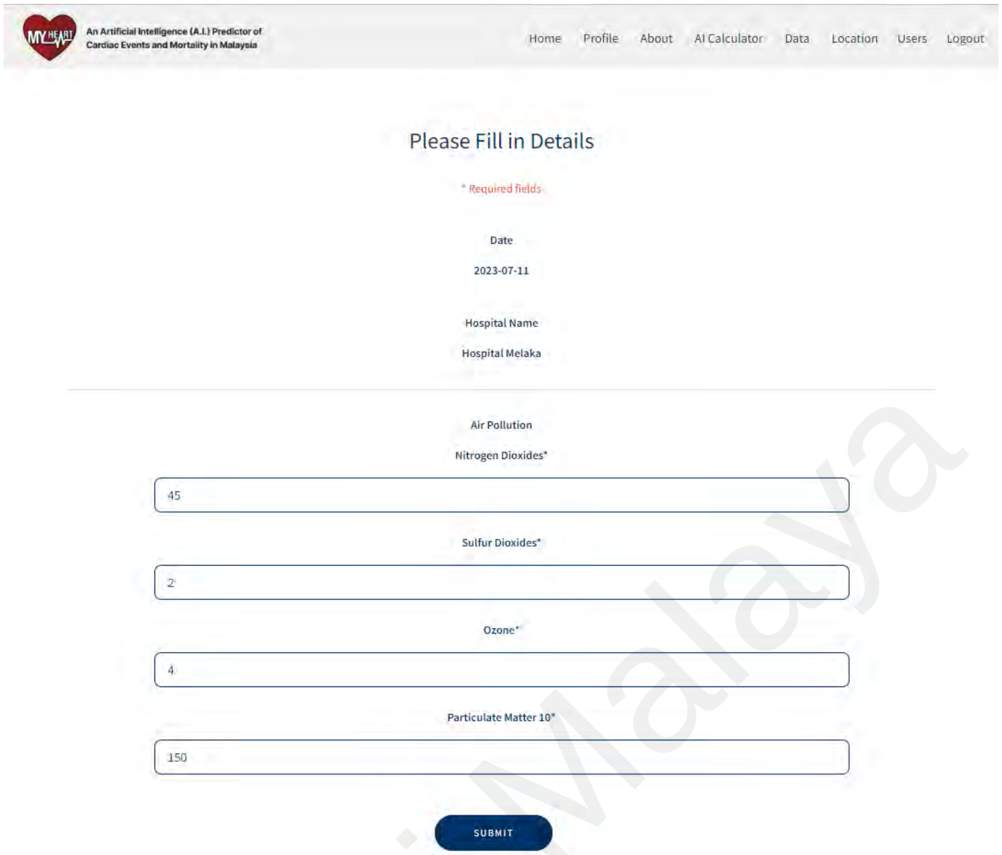
The input of the data is from the 'Hospital (Single) Geo-location Prediction' feature. This allows the system to constantly update air quality data for each location. Furthermore, users can edit, update, and delete the air pollution data as needed. Figure 4.35 presents the user interface of the ‘Hospital Location and Air Quality Database' page, and Figure 4.36 presents 'Edit Air Quality Data' page.


MyHeart - Machine Learning Co
myheartacsair.uitm.edu.my/sistem/HospitalPassRecord.php#edit_info
An Artificial Intelligence (A.I.) Predictor of Cardiac Events and Mortality in Malaysia
Home
Profile
About
AI Calculator
Data
Location
Users
Logout
ADD DATA
TOP

Information of Hospital Location and Air Quality

SDP ID	Date	Hospital Name	Nitrogen Dioxides	Sulfur Dioxides	Ozone	PM10	Actions
1017	2023-07-11	Hospital Melaka	45	2	4	150	EDIT DELETE
1009	2023-07-13	Hospital Sultanah Bahiyah	23	4	1	154	EDIT DELETE
1001	2023-07-09	Pusat Perubatan Universiti Malaya	45	2	4	157	EDIT DELETE
1016	0000-00-00	Hospital Tengku Ampuan Afzan	30	17	9	89	EDIT DELETE
1001	2023-07-10	Pusat Perubatan Universiti Malaya	56	5	6	150	EDIT DELETE

Figure 4.35: The 'Hospital Location and Air Quality Database' page, displaying location-specific air quality readings.


MY HEART
An Artificial Intelligence (A.I.) Predictor of Cardiac Events and Mortality in Malaysia
Home
Profile
About
AI Calculator
Data
Location
Users
Logout
SUBMIT

Please Fill in Details

* Required fields

Date

2023-07-11

Hospital Name

Hospital Melaka

Air Pollution

Nitrogen Dioxides*

45

Sulfur Dioxides*

2

Ozone*

4

Particulate Matter 10*

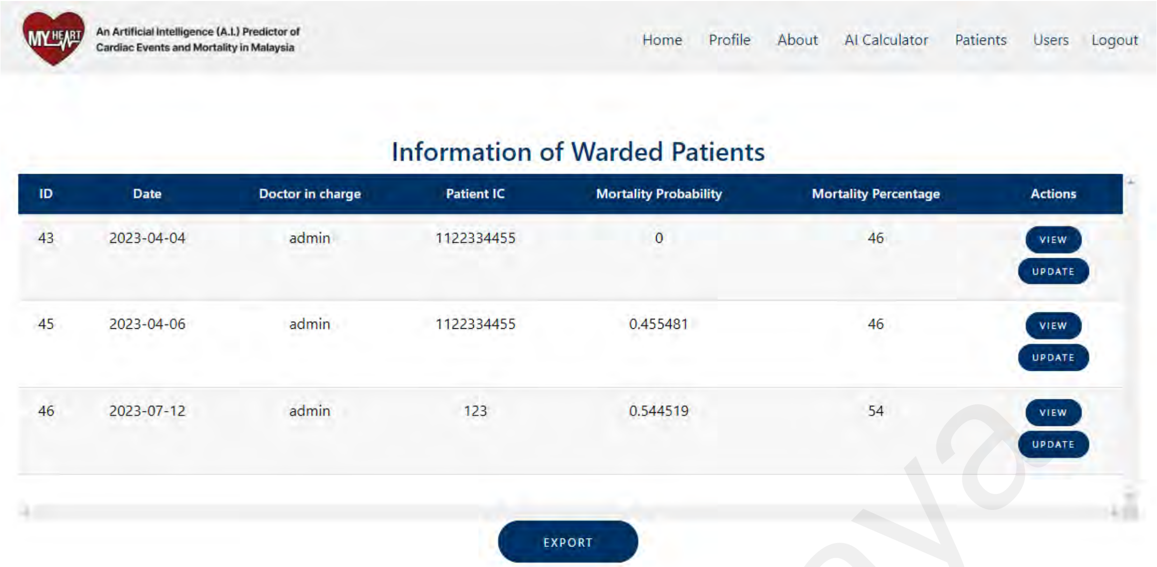
150

Figure 4.36: The 'Edit Air Quality Data' page, where users can update air quality readings in the database.

4.3.1.6.3 In-hospital ACS Patients Database

The ‘In-hospital ACS Patients Database’ serves as the data repository for hospitalized ACS patients. This database contains hospitalized patient-specific data, acquired through the 'ACS Mortality Prediction (In-hospital)' feature. It features patient information, including ID, date, assigned doctor, patient IC, calculated mortality probability, and the mortality percentage (Figure 4.37). There is an ‘EXPORT’ button on the bottom of the table, which allows users to download the stored information in CSV format.

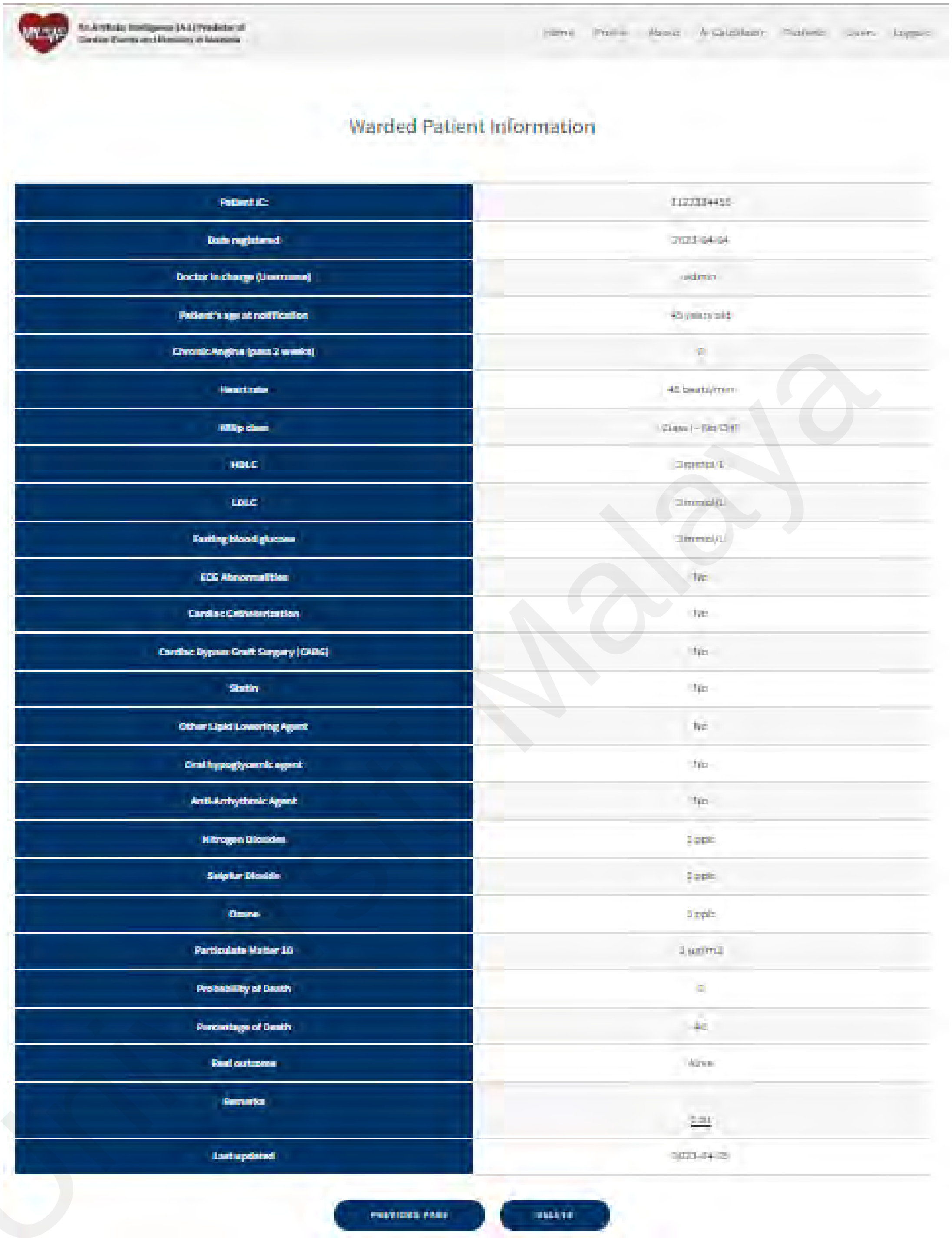
Besides, the database allows users to 'VIEW' and 'UPDATE' individual patient records. When users click on “VIEW”, the page will display the details of the patients’ health data as shown in Figure 4.38. When there are any changes to the patient’s status, the users can easily update the relevant patient record as well (Figure 4.39).


MY HEART
An Artificial Intelligence (A.I.) Predictor of Cardiac Events and Mortality in Malaysia
Home
Profile
About
AI Calculator
Patients
Users
Logout
VIEW
UPDATE
EXPORT

Information of Warded Patients

ID	Date	Doctor in charge	Patient IC	Mortality Probability	Mortality Percentage	Actions
43	2023-04-04	admin	1122334455	0	46	VIEW UPDATE
45	2023-04-06	admin	1122334455	0.455481	46	VIEW UPDATE
46	2023-07-12	admin	123	0.544519	54	VIEW UPDATE

Figure 4.37: User interface of the 'In-hospital ACS Patients Database', showing a summary of patient data and available user interactions, including the "VIEW", "UPDATE" and "EXPORT".

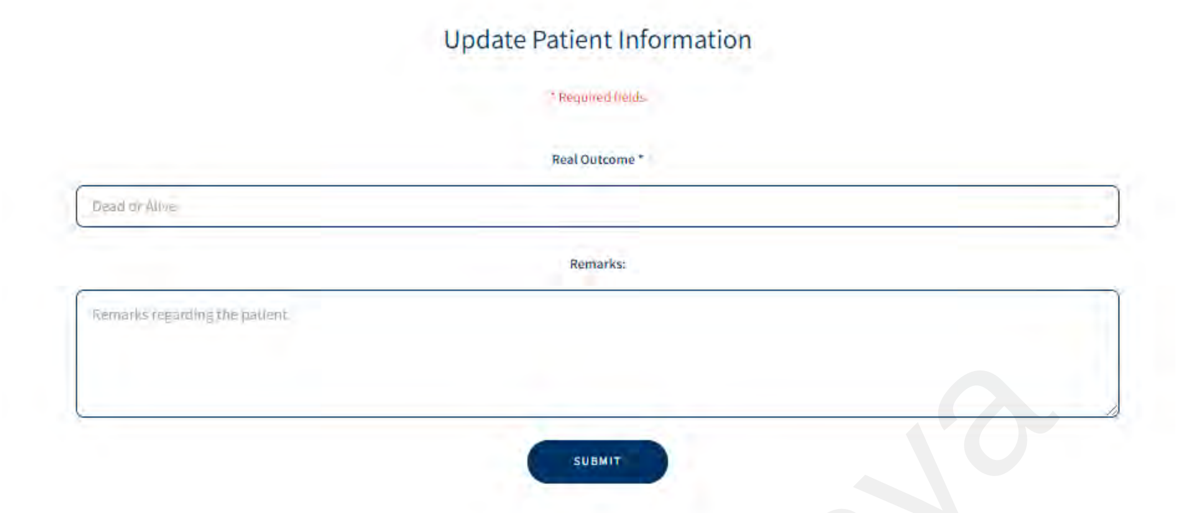


An Artificial Intelligence (A.I.) Predictor of Cardiac Events and Mortality in Malaysia
Home
Profile
About
AI Calculator
Patients
Users
Logout
PREVIOUS PAGE
DELETE

Warded Patient Information

Patient IC:	1122334455
Date registered	2023-04-04
Doctor in charge (Username)	admin
Patient's age at notification	45 years old
Chronic Angina (past 2 weeks)	0
Heart rate	45 beats/min
Killip class	Class I - No CHF
HDLC	3 mmol/L
LDLC	3 mmol/L
Fasting blood glucose	3 mmol/L
ECG Abnormalities	No
Cardiac Catheterization	No
Cardiac Bypass Graft Surgery (CABG)	No
Statin	No
Other Lipid Lowering Agent	No
Oral hypoglycaemic agent	No
Anti-Arrhythmic Agent	No
Nitrogen Dioxides	3 ppb
Sulphur Dioxide	3 ppb
Ozone	3 ppb
Particulate Matter 10	3 µg/m3
Probability of Death	0
Percentage of Death	4c
Real outcome	Alive
Remarks	Edit
Last updated	2023-04-05

Figure 4.38: Detailed patient record view within the 'In-hospital ACS Patients Database', displaying patient information.



Update Patient Information

* Required fields

Real Outcome *

Dead or Alive

Remarks:

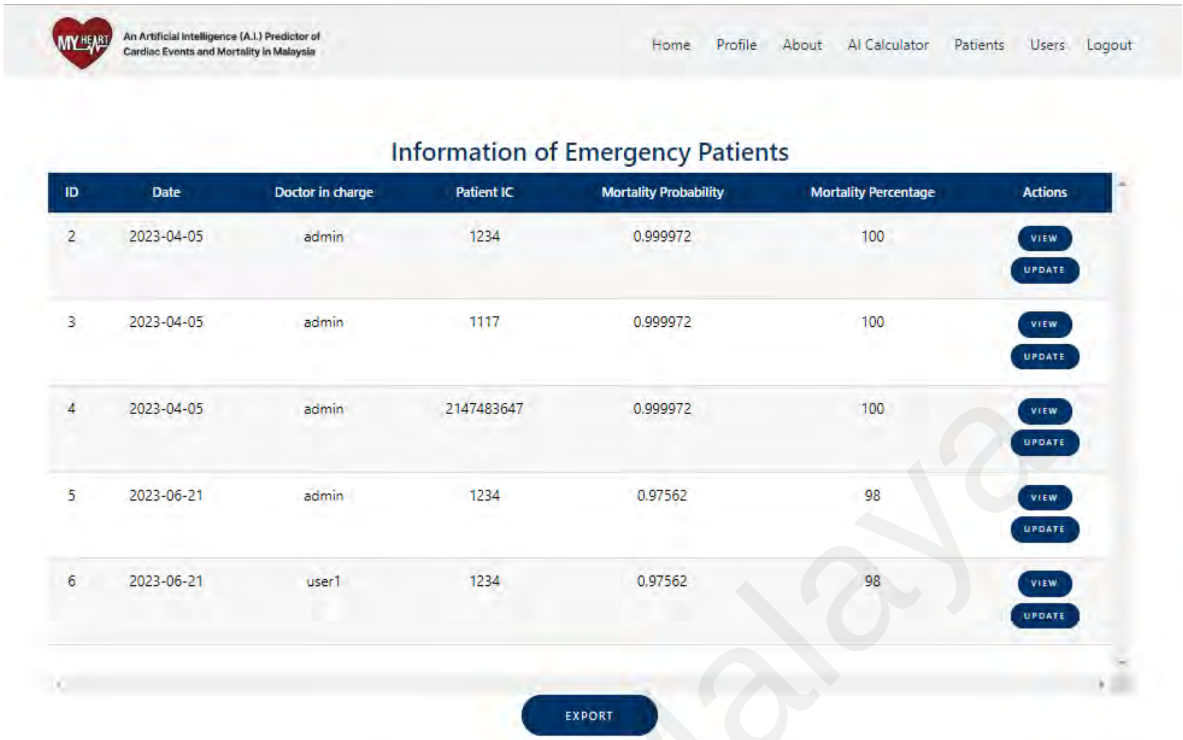
Remarks regarding the patient

SUBMIT

Figure 4.39: Update interface within the 'In-hospital ACS Patients Database', allowing updates and remarks for the specific patients as needed.

4.3.1.6.4 Emergency ACS Patients Database

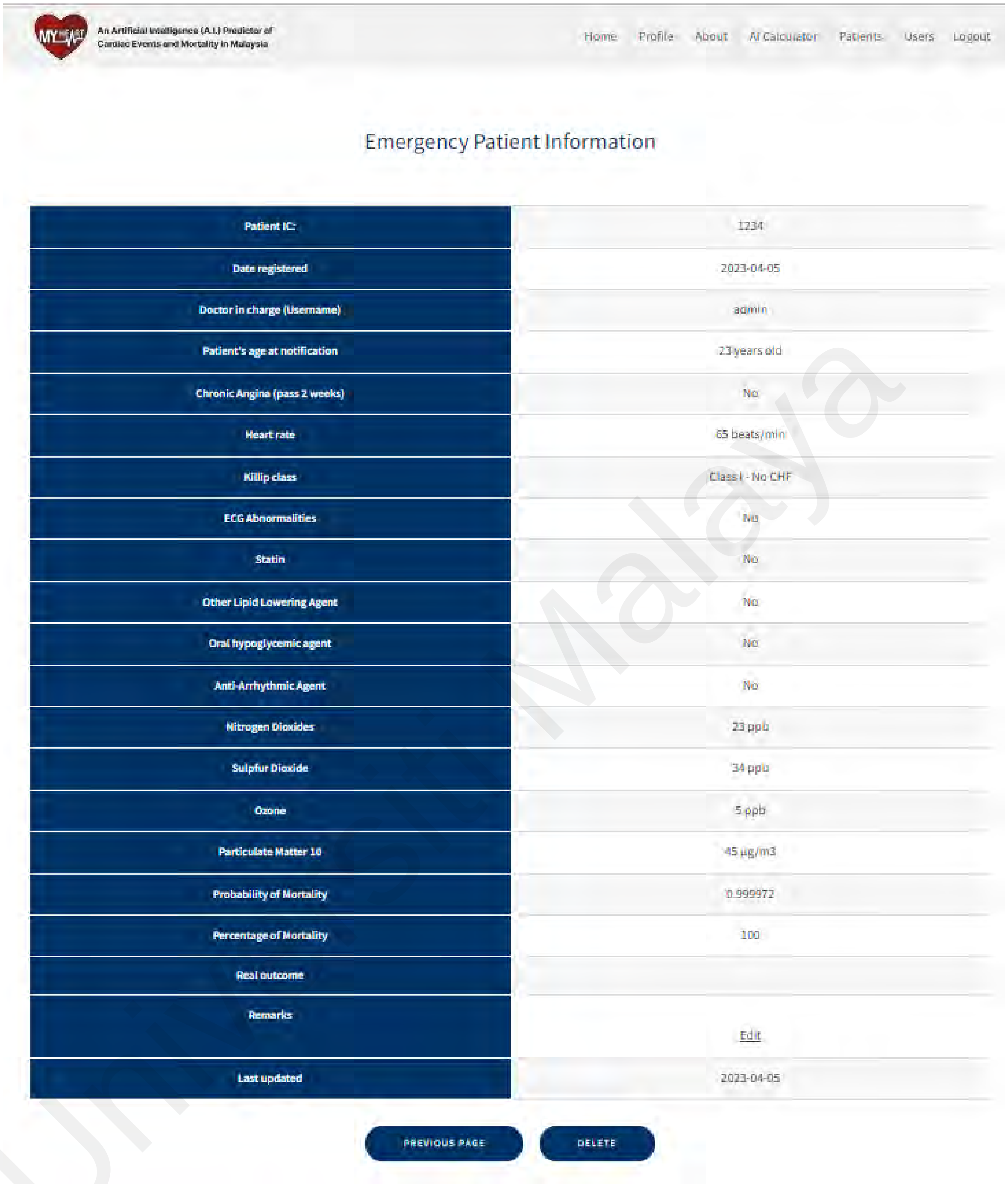
The 'Emergency ACS Patients Database' is a subsystem of the 'MyHeart ACS Air' website. Similar to the ‘In-hospital ACS Patients Database’, this database focuses on emergency ACS patients, where the data collected from 'ACS Mortality Prediction (Emergency)' feature. This database is essential for emergency patient data management for calculating risk predictions during emergency situations. Figures 4.40 – 4.42 show the screenshots of the emergency ACS patient database pages.


MY HEART
An Artificial Intelligence (A.I.) Predictor of Cardiac Events and Mortality in Malaysia
Home
Profile
About
AI Calculator
Patients
Users
Logout
EXPORT

Information of Emergency Patients

ID	Date	Doctor in charge	Patient IC	Mortality Probability	Mortality Percentage	Actions
2	2023-04-05	admin	1234	0.999972	100	VIEW UPDATE
3	2023-04-05	admin	1117	0.999972	100	VIEW UPDATE
4	2023-04-05	admin	2147483647	0.999972	100	VIEW UPDATE
5	2023-06-21	admin	1234	0.97562	98	VIEW UPDATE
6	2023-06-21	user1	1234	0.97562	98	VIEW UPDATE

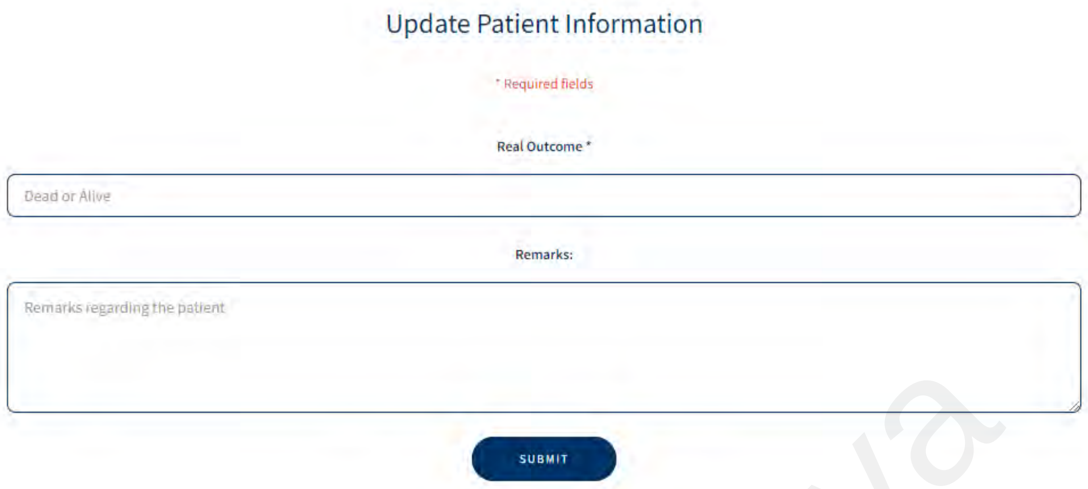
Figure 4.40: The 'Emergency ACS Patients Database' page, featuring an overview of emergency patient data and user-interaction options.


MY HEART
An Artificial Intelligence (A.I.) Predictor of Cardiac Events and Mortality in Malaysia
Home
Profile
About
AI Calculator
Patients
Users
Logout

Emergency Patient Information

Patient IC:	1234
Date registered	2023-04-05
Doctor in charge (Username)	admin
Patient's age at notification	23 years old
Chronic Angina (pass 2 weeks)	No
Heart rate	65 beats/min
Killip class	Class I - No CHF
ECG Abnormalities	No
Statin	No
Other Lipid Lowering Agent	No
Oral hypoglycemic agent	No
Anti-Arrhythmic Agent	No
Nitrogen Dioxides	23 ppb
Sulpfur Dioxide	34 ppb
Ozone	5 ppb
Particulate Matter 10	45 µg/m3
Probability of Mortality	0.999972
Percentage of Mortality	100
Real outcome	
Remarks	Edit
Last updated	2023-04-05

Figure 4.41: Detailed view of an individual patient record within the 'Emergency ACS Patients Database'.



Update Patient Information

* Required fields

Real Outcome *

Dead or Alive

Remarks:

Remarks regarding the patient

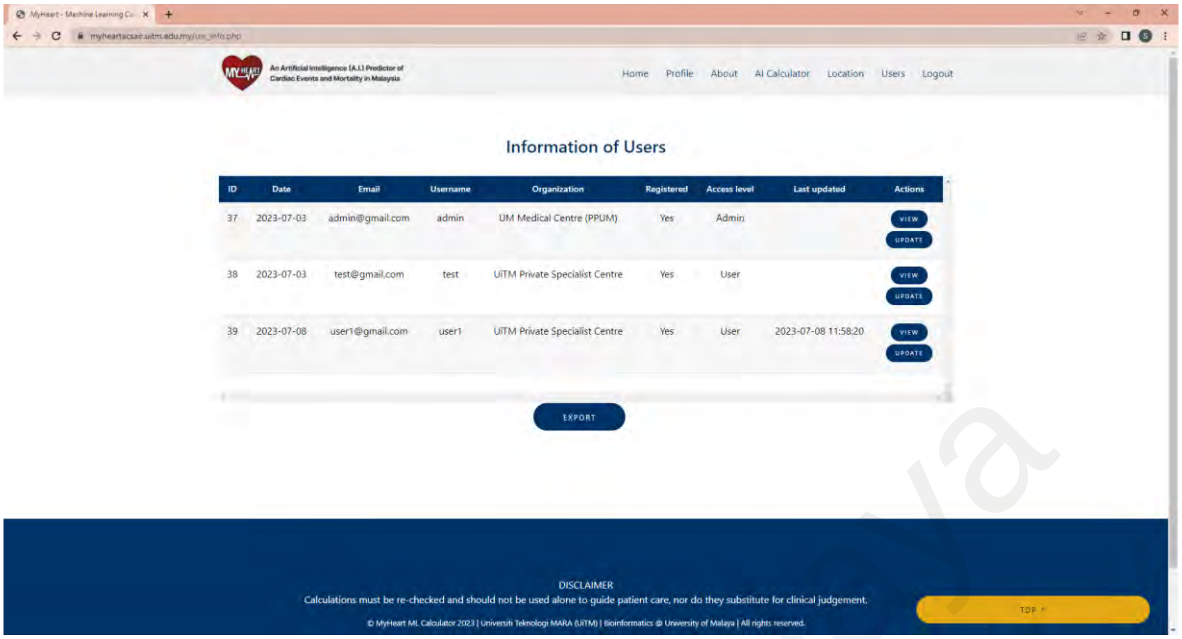
Figure 4.42: The 'Update Patient Information' interface within the 'Emergency ACS Patients Database', allows users to update and include remarks when necessary.

4.3.1.6.5 Users Database

The ‘Users Database’ maintains a record of all users registered on the ‘MyHeart ACS Air’ system (Figure 4.43). This database is exclusively accessible by the system administrator and contain user registered information including the user ID, registration data, email address, username, associated organization, registration status, access level, and last updated timestamp. The users’ passwords are encrypted and protected from the access of administrator.

Aside from providing an overview of user information, this database allows for interactive user management. In the 'Actions' column, the admin has options to 'VIEW' individual user profiles (Figure 4.44) and 'UPDATE' user details as required (Figure 4.45).

The 'VIEW' option opens a detailed page with a user's complete profile. On the other hand, the 'UPDATE' option takes the admin to a separate interface where changes to the user's profile, such as access level adjustments or account activation, can be made.


MyHeart - Machine Learning Ca
An Artificial Intelligence (A.I.) Predictor of Cardiac Events and Mortality in Malaysia
Home
Profile
About
AI Calculator
Location
Users
Logout
VIEW
UPDATE
EXPORT
TOP

Information of Users

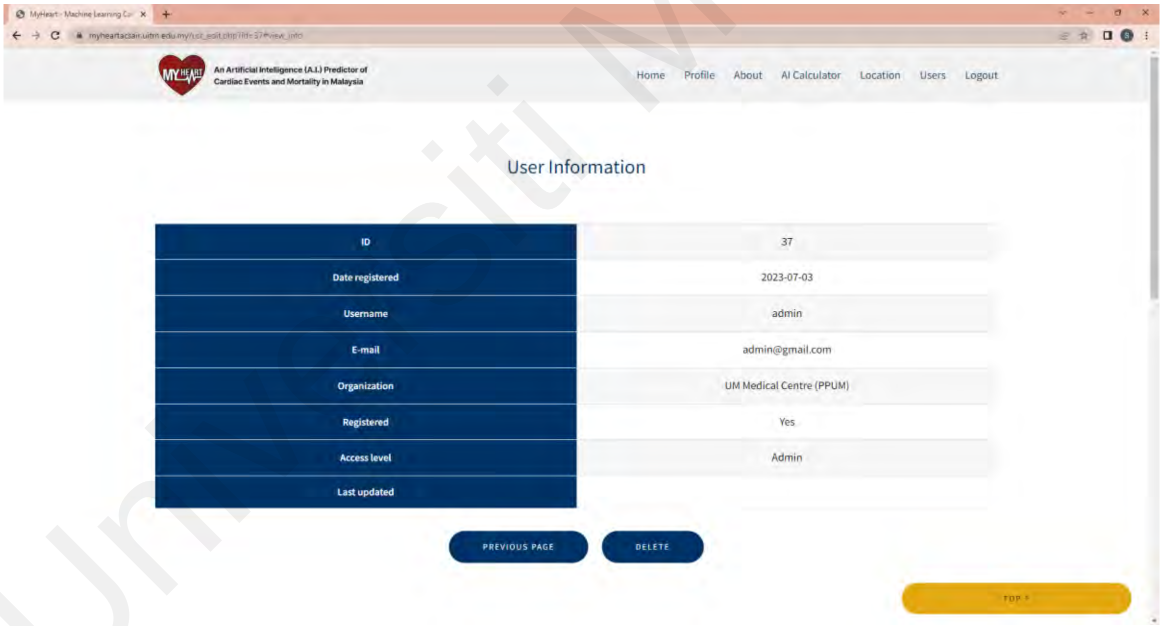
ID	Date	Email	Username	Organization	Registered	Access level	Last updated	Actions
37	2023-07-03	admin@gmail.com	admin	UM Medical Centre (PPUM)	Yes	Admin		VIEW UPDATE
38	2023-07-03	test@gmail.com	test	UiTM Private Specialist Centre	Yes	User		VIEW UPDATE
39	2023-07-08	user1@gmail.com	user1	UiTM Private Specialist Centre	Yes	User	2023-07-08 11:58:20	VIEW UPDATE

DISCLAIMER

Calculations must be re-checked and should not be used alone to guide patient care, nor do they substitute for clinical judgement.

© MyHeart ML Calculator 2023 | Universiti Teknologi MARA (UiTM) | Bioinformatics @ University of Malaya | All rights reserved.

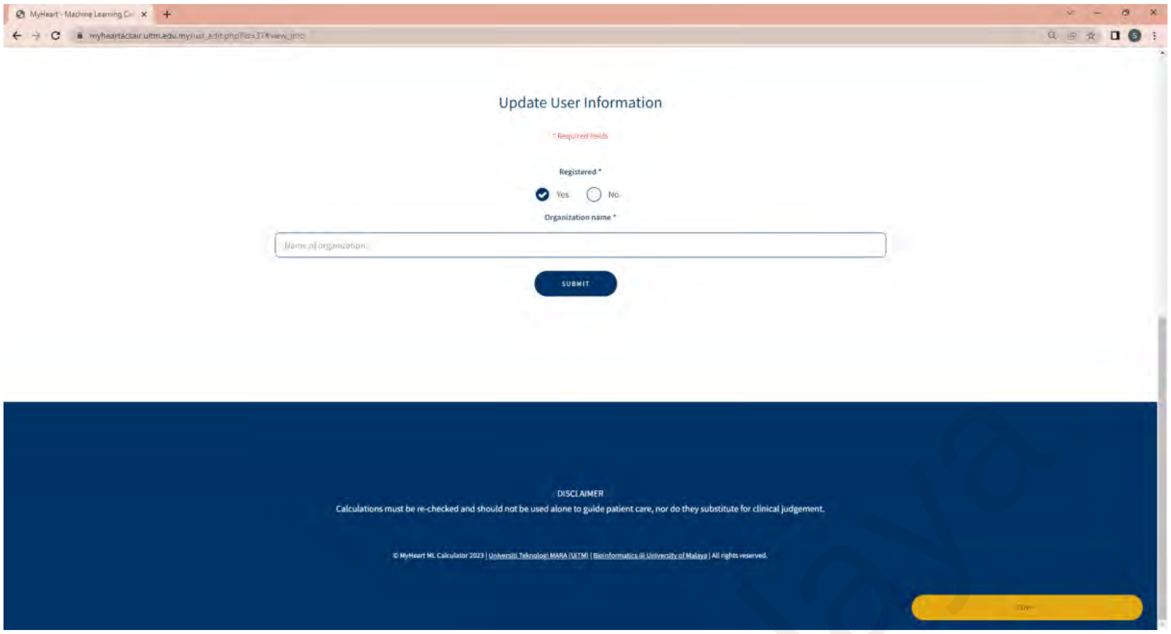
Figure 4.43: The 'Users Database' page, displaying an overview of registered user information and interactive management options.


MyHeart - Machine Learning Ca
An Artificial Intelligence (A.I.) Predictor of Cardiac Events and Mortality in Malaysia
Home
Profile
About
AI Calculator
Location
Users
Logout
PREVIOUS PAGE
DELETE
TOP

User Information

ID	37
Date registered	2023-07-03
Username	admin
E-mail	admin@gmail.com
Organization	UM Medical Centre (PPUM)
Registered	Yes
Access level	Admin
Last updated	

Figure 4.44: Detailed view of an individual user profile within the 'Users Database'.


MyHeart - Machine Learning Ca
myheartacsair.uitm.edu.my/usr_edit.php?id=37#view_info

Update User Information

* Required fields

Registered *

● Yes ○ No

Organization name *

Name of organization

DISCLAIMER
Calculations must be re-checked and should not be used alone to guide patient care, nor do they substitute for clinical judgement.

© MyHeart ML Calculator 2023 | Universiti Teknologi MARA (UiTM) | Bioinformatics @ University of Malaya | All rights reserved.

Figure 4.45: The 'Update User Information' interface within the 'Users Database', designed for simplified user administration and profile updates.

4.3.2 System Usability Scale (SUS)

The system user usability test evaluation form is created based on the System Usability Scale (SUS), developed by (Brooke, 1996), was employed to evaluate the usability of the ‘MyHeart ACS Air’ web system. The SUS is a reliable tool for measuring the usability and functionality of a website. It only comprises 10 questions with five response options from “Strongly Agree” to “Strongly Disagree”, each respond corresponds to a specific score, as shown in Table 4.12 below.

Table 4.12: System Usability Scale (SUS) score distribution.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

The system usability questionnaire is given to potential users of the website via Google Forms. These users mainly include medical personnel - especially cardiologists, as well as researchers. Upon deployment of the web system, we collect responses from the users and calculate scores based on the System Usability Scale (SUS) methodology. The 'MyHeart

ACS Air' web system achieved an average SUS score of 75. For more details on the SUS questionnaire, please refer to Appendix E.

The SUS survey consists of 10 questions, where each question contributes equally to the final SUS score. Each question is alternately positive and negative, to make sure that the users read through the questionnaire thoroughly. Question 1, 3, 5, 7, and 9 are positive questions, the higher the scores are better. As for question 2, 4, 6, 8, and 10 are negative questions, lower scores are better.

Table 4.13 presents the SUS detailed breakdown of SUS questions with the question type, the mean rating and percent agree, that gives clearer insights in which aspects of the system users find particularly usable or problematic.

Table 4.13: Detailed breakdown of SUS questions, question type, mean rating, and percentage agreement based on user responses.

SUS Questions	Question Type	Mean Rating	Percent Agree (%)
1. I think that I would like to use MyHeart ACS Air System frequently.	Positive	4.5	91
2. I found MyHeart ACS Air System unnecessarily complex.	Negative	2.5	18
3. I feel that MyHeart ACS Air System was easy to use	Positive	4.3	82
4. I think I would need the support of a technical person to be able to use MyHeart ACS Air System.	Negative	2.5	18
5. I found the various functions in MyHeart ACS Air System were well integrated.	Positive	4.5	91
6. I thought there was too much inconsistency in MyHeart ACS Air System.	Negative	2.0	0
7. I would imagine that most people would learn to use MyHeart ACS Air System very quickly.	Positive	4.3	82

Table 4.13, continued.

SUS Questions	Question Type	Mean Rating	Percent Agree (%)
8. I found MyHeart ACS Air System very cumbersome (awkward) to use.	Negative	1.7	0
9. I felt very confident using MyHeart ACS Air System.	Positive	4.2	91
10. I need to learn a lot of things before I could use MyHeart ACS Air System.	Negative	2.4	18

*Percent Agree (%) = Agree (4) and Strongly Agree (5) responses combined.

According to Bangor, et al. (2009), a SUS score of 68 or higher is considered above average and acceptable by the user, as illustrated in Figure 4.47 below. As a result, a score of 75 corresponds to a 'B' grade. This shows a high level of usability. The system is acceptable to users, and users evaluated the system to be user-friendly and useful based on its functionality, however the system can yet be enhanced. The SUS test collects user feedback and identifies areas where future system versions can seek to improve usability and user experience even further.

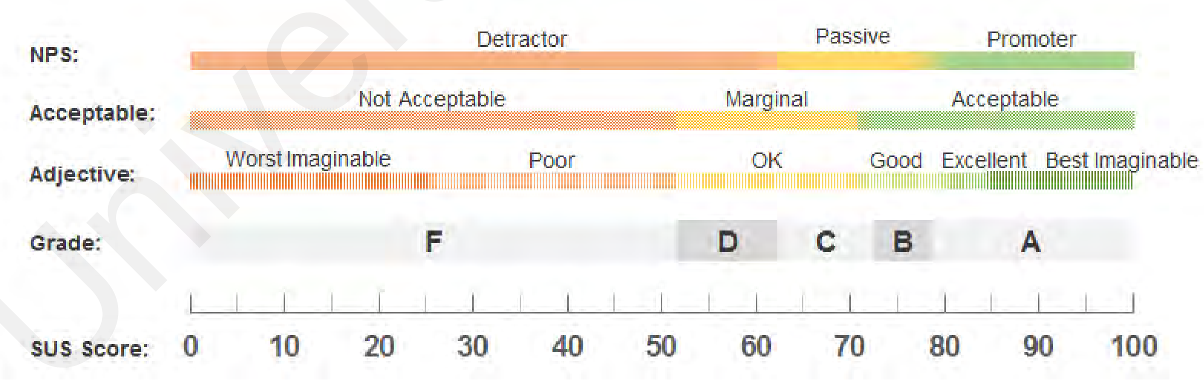


Figure 4.46: SUS scores grade rankings from “Determining what individual SUS scores mean: Adding an adjective rating scale.” (Photo sourced from Bangor, et al., 2009)

CHAPTER 5: DISCUSSIONS

Air pollution has been widely acknowledged as a significant health risk, especially for cardiovascular diseases (CVD) such as Acute Coronary Syndrome (ACS) (Zhao, et al., 2023; Rus & Mornoş, 2022; Kuźma, et al., 2019). However, most studies on this topic are limited in the context of Southeast Asia, particularly Malaysia (Liu, et al., 2022; Kuźma, et al., 2021; Santurtún, et al., 2017),

The database used for this study is unique in that it includes the three major ethnicities in Asia: Chinese, Indian, and Malay. Previous research relied on a homogeneous population database, raising concerns about its applicability to the Asian continent. Given the unique environmental and demographic characteristics of the country (Swee-Hock, 2015). Moreover, traditional risk scoring models (TIMI and GRACE) often overlook environmental factors (Antman, et al., 2000; Granger, et al., 2003), thus offering potential for improved predictive accuracy by incorporating these important variables.

This study addresses these gaps by using machine learning (ML) and stacked ensemble learning (EL) models to predict the (i) hospitalization, (ii) mortality rate of ACS and (iii) the ACS mortality risk in relation to air pollution in Malaysia guided by the objectives and research questions. This novel approach enables better predictive accuracy and understanding of the effects of various air pollutants on the incidence of ACS using ML, stacked EL and SHAP analysis.

The web system is developed by integrate the best performing ML models from this study. This system provides an interface where users can interact with the predictive models and visualize the predicted ACS hospitalization and mortality cases, displayed via Google Maps. Moreover, it encompasses a mortality risk calculator, thus allowing users to gain insight into the ACS mortality risk in the presence of air pollution.

This web system facilitates a better understanding of the impact of air pollution on ACS patients and demonstrate the significance of including environmental factors in risk assessment models. Besides, it also highlights the potential of ML and stacked EL in generating prediction that are more accurate and comprehensive. As a result, it contributes to research on the onset of ACS in Malaysia, potentially guiding more effective prevention and management strategies.

5.1 The Impact of Air Pollution on Acute Coronary Syndrome (ACS) Hospitalization and Mortality Rate: A Regression Analysis

These models utilized air pollution metrics as key predictive variables, examined over four specific time lags (00, 03, 07, and 30 days). Time lag 00 and time lag 03 are associated with short term exposure to patients, meanwhile time lag 07 and time lag 30 are associated with long term exposure (Bourdrel, et al., 2017). The choice of these time lags was guided by literature, which often reported immediate (lag 00) and short-term (lag 03) impacts of air pollution on ACS incidence (Zhao, et al., 2023; Liu, et al., 2022). This study also included longer average time lags of 07 and 30 days to capture potential weekly and monthly patterns in the relationship between air pollution and ACS events. Longer time lags produce models with higher RMSE value compared shorted time lag for prediction ACS hospitalization and mortality rate. The RMSE reported for time lags 03, 07 and 30 are higher, lower RMSE indicates better model performance (Ameer, et al., 2019). Conventional statistical methods, such as conditional logistic regression, are commonly used to investigate the effect of short-term air pollution exposure on ACS, but ML research on this topic is limited (Zhao, et al., 2023; Kranc, et al., 2021). As a result, this study focuses on short-term (time lag 00) air pollution exposure with ACS using ML for real-time predictions and web system integration. This study's discussion is based on short-term exposure to air pollution (time lag 00).

As a results, the findings for section 5.1 presents several insights based on time lag 00: (i) In comparison to time lags 03, 07, and 30, time lag 00 demonstrated the best ML predictive performance for ACS hospitalization and mortality based on air pollution features, (ii) ML outperformed EL in predicting ACS hospitalization and mortality rate, (iii) The RF model outperforms the other ML models in terms of predictive performance for ACS hospitalization, with an RMSE of 1.701, (iv) XGBoost model demonstrated better performance for ACS mortality rate prediction with the RMSE of 0.440, and (v) the SHAP summary plot indicated that nitrogen oxides (NOx) and ozone (O_3) is associated with increase of hospitalization and mortality rate in ACS patients.

The RMSE of 1.701 for ACS hospitalizations using RF model suggests an average prediction deviation of 1 to 2 hospitalized patients per day. Meanwhile, the RMSE of 0.440 for the predicted ACS mortality rate by the XGBoost model indicates that predictions typically deviate by 0 to 1 predicted mortality rate.

The ability of the RF algorithm to capture complex nonlinear relationships and handle large datasets with higher dimensionality contributed to its better performance in this study. RF is less sensitive to noise and outliers in the data and generates mean prediction of the individual tree derived from a large number of decision trees (Breiman L., 2001).

The XGBoost model demonstrated to be the best algorithm for predicting ACS mortality rate. XGBoost is an optimized gradient-boosting ML algorithm, it can effectively capture minor variances in the data and provide more accurate predictions by refining its predictions through multiple iterations especially for constrained ranges (Chen, et. al., 2016).

While all the developed models performed similarly in predicting ACS-related hospitalizations and mortality rates at lag 0, the RF and XGBoost models performed better, which agrees with existing literature that uses ML in hospitalization and mortality rate predictions (Kim, et al., 2022; Angraal, et al., 2020; Goto, et al., 2019).

In this study, stacked EL is used to improve predictive performance by integrating the base ML models (Jason, 2021). Despite the fact that a stacking ensemble model with GLM as the meta learner did not show a significant improvement in RMSE (ACS Hospitalization = 1.922; ACS Mortality = 0.444) when compared to the individual models in our study. This could be attributed to the individual base models' competent performance in predicting outcomes, resulting in limited improvement through stacked EL (Kalcheva, et al., 2020).

ML models are frequently regarded as black-box models. SHapley Additive exPlanations (SHAP) is a novel methodology for examining the influence of predictor variables and their interactions in ML models and breaking the "black box" paradigm that underpins the application of automatic ML techniques (Lundberg & Lee, 2017). The SHAP summary plots in Figures 4.7(a) and 4.7(b) were used to interpret the ACS hospitalization and mortality rate based on air pollutants such as nitrogen oxides (NOx), sulphur dioxides (SO_2), ozone (O_3), and particulate matter 10 (PM10) much easier and clearer in this study.

According to the SHAP summary plots, NOx and O_3 were the top two contributors. A high concentration (shown in red) of these pollutants in the plot's positive SHAP value area. This implies that high levels of NOx and O_3 in the atmosphere significantly increase the risk of hospitalization and mortality among ACS patients, providing clear evidence of their negative health effects, which is consistent with previous research on the negative health effects of these pollutants (Zhao, et al., 2023; Cheng, et al., 2020; Butland, et al., 2016; Raza, et al., 2014).

In Rus & Mornoş (2022) study on the pathophysiological mechanisms of air pollutants, stated that NOx contributes to endothelial dysfunction, as well as prothrombotic and proinflammatory effects, which can lead to ACS. Furthermore, researchers discovered there is a correlation between NO_2 exposure and the incidence of NSTEMI (Butland, et al., 2016; Wang, et al., 2015). This study aligns with Jiang, et al. (2023) study, affirming that acute exposure of O_3 was associated with increased cardiac hospitalization, however, the exact mechanism by which ozone affects ACS patients is not fully understood, since most of the existing literature addressed the statistical correlation, leaving the biological pathways unclear.

In contrast, the SHAP summary plots of SO_2 and PM10 revealed no significant influence of these pollutants on the outcomes, exhibiting a near-central tendency with no significant deviation. This observation presents an intriguing contrast to some existing literature that suggests potential adverse health effects of SO_2 and PM10 (Díaz-Chirón, et al., 2021; Kuźma, et al., 2019; Zhao, et al., 2016; Lippi, et al., 2010).

Using the SHAP summary plots, this study was able to identify and isolate the effects of each air pollutant, providing a clearer understanding of each feature's contribution in ACS events.

5.2 Predicting Mortality Risk in Patients with ACS in the Context of Air Pollution: A Classification Machine Learning (ML) Approach

To the best of our knowledge, no studies that incorporate environmental factors into a mortality risk prediction model have been conducted. This is the first study to show that in-hospital and emergency mortality in Malaysian patients with ACS is predicted with air pollution. To predict the mortality risk of patients with ACS in Malaysia, multivariate clinical features with air pollution features were used to develop ML models with stacked EL. The ML and stacked EL models were also validated using traditional risk scores (TIMI).

The findings of this study can be summarised as follows: i) RF (AUC = 0.840) outperform other ML and EL models when using in-hospital selected features. ii) In emergency selected features dataset, the XGBoost algorithm outperforms other ML and EL models with the highest AUC (AUC = 0.844). iii) Both ML model and stacked EL developed using in-hospital and emergency features (AUC ranging from 0.82 – 0.84) outperformed conventional risk scoring score TIMI in in-hospital features (STEMI AUC = 0.791 and NSTEMI AUC = 0.659) and TIMI in emergency features (STEMI AUC = 0.797 and NSTEMI AUC = 0.659) iv) SHAP summary plots illustrate the model's explainability, among the air pollutants, NOx and O_3 shows impacts towards the mortality risk in ACS patients.

Previous research has shown that models based on ML perform better in classification tasks than models based on conventional risk scores in ACS mortality studies (Kasim et al., 2022a; Kasim et al., 2022b; Ke et al., 2022; Wu et al., 2021; Aziz et al., 2021; Aziida et al., 2021; Aziz F. et al., 2019).Similar findings were reported in our study as well; this study introduces a novel approach by integrating environmental factors, specifically air pollution features, with clinical features to enhance mortality risk prediction using ML and stacked EL approach. The absence of environmental factors in conventional risk scoring method is notable, given the growing evidence of the influence of environmental factors, specifically air pollution, on cardiovascular health (Pope, et al., 2011).

In this study the best ML prediction model, RF resulted in an AUC of 0.843 (95% CI: 0.813 – 0.873) for STEMI and 0.842 for NSTEMI (95% CI: 0.795 – 0.889), based on the in-hospital selected features dataset. Meanwhile the TIMI risk score achieved an AUC of 0.791 (95% CI: 0.757 – 0.825) for STEMI and 0.565 (95% CI: 0.505 – 0.625) for NSTEMI.

In the emergency selected features dataset, the best performing ML prediction model, XGBoost yielded an AUC of 0.841 (95% CI: 0.821 – 0.862) for STEMI and 0.84 (95% CI: 0.809 – 0.872) for NSTEMI. The TIMI risk score obtained an AUC of 0.797 (95% CI: 0.774 – 0.82) for STEMI and 0.583 (95% CI: 0.543 – 0.622) for NSTEMI.

Application of ML algorithms is promising for predicting the in-hospital and emergency mortality of ACS patients in the presence of air pollution, particularly the RF algorithm and XGBoost algorithm that exhibited superior performance. According to Liaw & Wiener (2002), the RF algorithm is known for its robustness in handling high-dimensional data and complex inter-feature interactions. In VanHouten, et al. (2014) study, RF model (AUC = 0.848) outperforms elastic net, ridge regression, and conventional TIMI and GRACE risk scores in predicting ACS mortality risk, which are similar to the classification model findings in predicting in-hospital mortality risk of ACS patients in this study.

The XGBoost (AUC = 0.958 [95% CI: 0.938 – 0.978]) showed promising performance in predicting mortality risk in patients with ACS (Wu, et al., 2021). In Ke, et al. (2022) study aimed to identify in-hospital mortality risk factors in ACS patients and compare the performance of ML prediction models. The XGBoost has the highest AUC value (AUC = 0.918) among all other predictive models including RF (AUC = 0.913), logistic regression (AUC = 0.884) and SVM (AUC = 0.896). The key risk factors identified included NT-proBNP, D-dimer, and Killip class. However, the study does not include air pollution features.

The reason for the high performance of the XGBoost models can be explain by its gradient boosting mechanism, which enhances the predictions gradually, the algorithm generated a series of decision trees in a gradient boosting manner, and produced the next decision tree based on the current one to better predict the outcome (Chen & Guestrin, 2016). This feature is especially suited to the less complex, streamlined emergency dataset, where iterative error correction can lead to highly accurate predictions.

Stacked EL was also employed in this study to potentially enhance the performance of the ML models. However, given the robust performance of our base models, the EL did not demonstrate significant improvements. This is consistent with previous research by Zhang, et al. (2022), in which stacked EL provided limited improvement when base models already provided higher predictive value, owing to its complexity in model interpretation. A recent study by Kasim S, et al. (2023) focused on predicting in-hospital mortality in Asian women post-STEMI using ML and stacked EL using the same NCVD dataset and the models were compared to the conventional TIMI risk score, proven that ML and EL techniques provided more accurate classifications for Asian women with STEMI than traditional methods. SVM Linear, an individual ML model, outperformed the best stacked EL model.

The TIMI score's simplicity is recognized in current guidelines and is frequently used in Asia hospitals for risk assessment of patients with ACS. The TIMI risk score, originally established to predict mortality outcomes, its application has since been extended and it is widely employed to predict various mortality post-ACS onset (Chimparlee, et al., 2018; Timbol, et al., 2015; Correia, et al., 2014; González-Pacheco, et al., 2012; Ahmad, et al., 2011). It was reported that the TIMI score is better than GRACE score calibration because it has more variables associated with ACS mortality, a balanced distribution of low, intermediate, and high-risk patients, and more accurate estimation (Lee, et al., 2018). In a comparative analysis with the widely accepted TIMI risk score, ML models in this study demonstrated improved predictive performance compared to TIMI risk score, especially for NSTEMI patients.

Even though the TIMI risk score is widely used in the Asian population, it was developed using data from a Western Caucasian cohort with limited data from an Asian population. A previous validation study in the Asian population reported a modest accuracy for risk prediction for TIMI risk score in STEMI with an AUC of 0.78 (Selvarajah, et al., 2013). Other conventional risk scores also performed modestly when validated in Korean registry study for STEMI and NSTEMI patients using AUC as a performance metric GRACE (0.851 0.810), ACTION (0.852, 0.806) and TIMI score (0.781, 0.593) (Lee, et al., 2021). Similar moderate results of TIMI risk score were also demonstrated in this study, the TIMI score had a validation performance of 0.83 for STEMI and 0.55 for NSTEMI using the in-hospital dataset, and 0.79 for STEMI and 0.59 for NSTEMI using the emergency dataset.

This is further supported from the findings from net reclassification improvement (NRI) of STEMI and NSTEMI patients using the in-hospital selected variables produced a NRI of 8.71%, and 86.94% respectively when compared to the original TIMI risk score. As for emergency selected variables shows an improvement of 5.95% for STEMI and 50.75% for NSTEMI in the context of air pollution. Despite its low NRI value for STEMI patients, we can see that significant improvement is added to the NSTEMI population, a cohort that accounts for half or more of all ACS cases worldwide. In medical field, a small increase in the performance of predictive models is vital and capable of giving a significant impact (Alahmar, Mohammed, & Benlamri, 2018). In this study, we found that TIMI underestimated mortality risk in both lower and higher risk groups. This may cause treatment to be delayed, increasing avoidable deaths.

The TIMI score has several notable limitations. First, TIMI was developed using data from fibrinolytic-eligible patients with STEMI where reperfusion therapy and drug-eluting stents were not regular treatment (Morrow, et al., 2000). Stains and antiplatelet medicines like prasugrel and ticagrelor are now part of our daily routine. Because TIMI risk scores only reflect the key prognostic indicators, valuable information maybe missed (Kwon, et al., 2019). Exclusion of high-risk patients is also another limitation of the risk score (Chen, et al., 2018). The TIMI risk score lacks risk factors associated with environmental health, specifically, it does not consider the impact of air pollution, which is increasingly recognised as a significant contributing factor in the health risks associated with ACS.

Also, the Asian cohort was found to be carrying an overall higher disease burden and risk compared to TIMI cohort. The situation is worsened by the environmental factors that are commonly found in the Asia, as studies have shown that Asian countries are heavily impacted by air pollution and significantly contributes to premature mortality (Lelieveld, et al., 2015; Kan, et al., 2012; Gurjar, et al., 2010).

The lack of assessment for the risk factors, reduced the TIMI risk score discriminatory performance (Feder, et al., 2015; Bawamia, et al., 2013). In addition, there are different scoring systems for STEMI cases and NSTEMI cases. The conventional TIMI score requires two distinct scores; TIMI for STEMI 8 risk factors include age, systolic blood pressure, heart rate, Killip class, anterior or left bundle infarction, prior history of angina, diabetes, or hypertension, and weight. Meanwhile, the TIMI Risk Score for patients with UA or NSTEMI is composed of seven equally weighted, binary variables (Aragam, et al., 2009). Age, aspirin use during the previous seven days, coronary artery disease (CAD) risk factors, known CAD, recent anginal episodes; ST-segment alterations of at least 0.5mm on the ECG at the time of initial presentation, and elevation of serum cardiac markers (Feder, et al., 2015). Kasim et al. (2022) have successfully identified 14 risk factors pertinent to mortality in Asian ACS patients, outperforming models developed via traditional statistical approaches.

The 14 features in ascending order based on the outcome in optimum AUC starting with Killip class, fasting blood glucose (FBG), heart rate, age, low density lipoprotein (LDL-C), oral hypoglycaemic agent, cardiac catheterization (CA), high density lipoprotein (HDL-C), antiarrhythmic agent, statin, chronic angina past 2 weeks, lipid lowering agent, ST-segment elevation ≥ 1mm in ≥ 2 contiguous limb leads and lastly is coronary artery bypass grafting (CABG), were subsequently included in our dataset, were subsequently included in our dataset and combined with the air pollution features based on the day of ACS onset. As for the emergency dataset, it is identical to the in-hospital dataset with lesser features, features that are excluded are baseline investigations (HDL-C, LDL-C, and FBG) and invasive therapeutic procedures (CA and CABG), the emergency features consist of a reduced set of variables that are easily accessible in emergency situations determined by the cardiologist, without the need for extensive testing or patient history. When integrated with air pollution parameters, these risk factors enhance the predictive accuracy of the ML models in discerning the impact of air pollution on ACS patients.

The SHAP allows us to understand and make logical inferences about how these variables were chosen as well as their impact on outcomes of for the best model. According to the SHAP summary plot (Figure 4.11 and Figure 4.12), patients with higher feature values of Killip class, fasting blood glucose, age, and heart rate all are associated with poorer outcome or non-survival, where similar findings are reported in literature (Van Den Berg & Body, 2018; Tang, et al., 2007). Statin and Oral hypoglycaemic medications also contribute significantly to managing patients with ACS, this finding also reflected in our SHAP analysis, where these pharmacological medications emerged as top features in determining mortality risk. Studies suggested that patients with ACS who took statins demonstrated a lower risk of subsequent cardiovascular events and mortality. For instance, in a research study by Sposito

& Chapman (2002), it was revealed that early initiation of statin therapy in ACS patients after an acute coronary event was associated with improved clinical outcomes.

In ACS patients with concurrent diabetes, oral hypoglycaemic medications are crucial for achieving good glycaemic control, which is associated with improved outcomes in ACS (Prattichizzo, et al., 2020). In line with this, the in-hospital dataset identified Fasting Blood Glucose (FBG) as a significant risk factor for ACS mortality. This is reflective in the SHAP analysis where statins and oral hypoglycaemic medication emerged as a top feature in predicting mortality risk, underlining their critical role in the management and prognosis of ACS patients.

As for air pollution association with the patient mortality, higher feature value of NOx and O_3 also contributes the ACS mortality risk. This findings from SHAP analysis corresponds to the baseline characteristics derived from conventional statistics in Table 4.4 indicates that there is significant association between mortality risk and these significant variables. There was a clear association observed between NOx and O_3 and the probability of mortality in patients with ACS. From the plots we can see that in red represent high value of NOx and O_3 contributes to the risk of mortality clearly. The results of this study align with the observation made in section 5.1, which identified NOx and O_3 as important factors that have impacts on the mortality risk of ACS patients, and on the rates of ACS hospitalizations and mortality and may trigger the onset of ACS. Given these results also similar with in Zhao, et al. (2023) study, stating that positive association between NO_2 and ACS patients, particularly patients with NSTEMI (Butland, et al., 2016). A study in China investigated the impact of six major air pollutants on CVD. The COX proportional hazards model showed that these pollutants had the greatest short-term effects, especially on the first day of exposure, notably, PM2.5, PM10, NO_2, and CO air pollutants. Given these consistent results, there's a compelling case for intensified mitigation strategies, specifically targeting reductions in NOx and O_3 emissions. Such measures could be instrumental in lowering the associated mortality risks for ACS patients.

In the broader landscape of research into the effects of air pollutants on ACS patients, much of the existing literature has underscored the impact of PM2.5 (Chen, et al., 2022; Zhao, et al., 2016; Meng X., et al., 2016). However, this study brings into focus the significant influence of NOx and ozone O_3. While PM2.5 remains a crucial focal point in many studies due to its known adverse health effects, our findings emphasize that other pollutants, specifically NOx and O_3, also warrant considerable attention, especially in areas where their concentrations are particularly high or on the rise.

The findings of this study are novel because this indicates the potential importance of including environment factors, which have been overlooked in conventionally risk assessment models. Although there is limited number of studies that have integrated environmental factors with conventional clinical features, our findings enhance the current understanding of ACS by providing a more comprehensive examination of the risk factors.

5.3 Web System Development and Evaluation

The web system was named "My Heart ACS Air" has been developed that uses ML to predict ACS events and mortality risk in Malaysia while emphasizing the impact of air pollution on the ACS cohort. The web system enables users to generate predictions, visualize data, and manage ACS patients in relation to Malaysian air pollution.

The prototyping method was utilised during the development of the 'My Heart ACS Air' system. The iterative approach started with the initial set up of the web system on a local server. Subsequently, the primary users were engaged in the evaluation and enhancement of the system, offering feedback and suggestions regarding its usability and functionality. The website had modifications and refinements based on the insights acquired from these sessions. The iterative process of testing and refinement continued until the system achieved user expectations and fulfilled the necessary requirements. The prototyping approach ensured that the end-product was both functional and user-centric.

The user interface was designed taking into consideration the wide range of potential users, including medical professionals, cardiologist, nurses, general practitioners, researchers, and possibly government policymakers. The My Heart ACS Air system's user interface is designed based on Schneiderman's Eight Golden Rules ensuring the efficiency and user-friendliness of the system (Shneiderman, 1986).

According to the Schneiderman's Eight Golden Rules of Interface Design, the system maintain consistency in the design elements with nice colour scheme and provides navigation bar for users to enhance overall efficiency. Besides, the system also provides informative feedback when engaging with the system, especially the ACS mortality calculator, after users provides the necessary input, if the patients are considered as high-risk, it will return "HIGH RISK PATIENT", informing the user that the patient required attention and extra treatment/care.

Furthermore, the interface has been designed to provides users with a sense of completion with every action sequence, while simultaneously reducing the potential errors through input validation, where each input is a required question, to ensure the user fill up all the input space, and for 'Yes' and 'No' question, the use radio buttons, that minimize error in input. Feedback messages to the user was also integrated to the system for example, once the user fill out the location information form and click the "Submit" button, the user will be directed to a new page stating that the data has been "Data Successfully Added".

The platform’s designed is kept simple with clear instructions, thereby ensuring that users consistently experience a sense of control. Lastly, the system is optimized to reduce cognitive and memory load by keeping the data entry minimal, as evidenced by its performance in website speed and page insight tests. Detailed results from these tests can be found in the Appendix H and Appendix I.

The System Usability Scale (SUS) was used as an assessment to evaluate the usability of "My Heart ACS Air”. The SUS was originally developed by Brooke (1996), it offers a reliable, yet 'quick and dirty', tool for assessing the usability of a system and has been widely use in usability testing (Grier, et al., 2013). The SUS focuses on the effectiveness, efficiency, and satisfaction of the user's experience, making it a crucial instrument for the post-development phase. The SUS Questionnaire was distributed using Google Form and feedback was sought from a diverse set of users, including cardiologists, nurses, and researchers as list out in Appendix F. Thus, the SUS matrix is capable of identifying potential areas for improvement by utilising user feedback. The feedback was received on unexpected system behaviours and error handling, which often neglected by the developers.

The system achieved a score of 77, equivalent to a grade of 'B’. Within the context of SUS evaluations, this score is indicative of a 'Good' user experience. This demonstrates the effectiveness of the iterative and prototype-driven development approach, which prioritised user feedback. Furthermore, the score reaffirms the system's high usability and aligns it with platforms that meet general user acceptability standards.

In essence, the SUS results indicate 'My Heart ACS Air's' success in achieving its objective: developing a web system that incorporates ML and visualization elements to provide users with a better understanding of the relationship between air pollution and ACS in Malaysia.

5.4 Significant of the Study

This study evaluating the impact of air pollution on the incidence of ACS using advanced ML which are integrated into a web-based system. This enables a wider user base to interact with and utilize the predictive model effectively. It provides a more holistic view of the potential triggers for ACS, which has not been traditionally considered. This study not only improves the accuracy of the risk assessment but also contributes to the field by highlighting the potential influence of environmental factors on health outcomes. Significantly, the study focuses on the Malaysian context, considering the unique geographical, environmental, and demographic factors of the region thereby enhancing its relevance and applicability for local healthcare providers and policymakers.

In considering the implications of our research, several significant benefits were highlighted as follows:

(i) Comprehensive Risk Assessment and Improved Prediction Models

This prospective study provides a comprehensive understanding of ACS cases by incorporating environmental risk factors, typically overlooked in risk assessments, to evaluate the risk of mortality and hospitalization in ACS patients. This method allows for the improvement of prediction, prevention, and management strategies.

From a clinical perspective, this research provides significant insights into variable factors affecting ACS mortality, presenting potential therapeutic targets. In terms of methodology, it demonstrates the application of ML predictive algorithms for healthcare professionals, particularly cardiologists and healthcare planner.

The ML models employed in this study offer more accurate and reliable predictions than the TIMI risk score. As such, the ML models algorithms outperform conventional risk models (Gibson, et al., 2020), which has significant implications for risk stratification and treatment, potentially improve measures and management strategies in ACS patients.

Even though the ML models utilised in this study are quite effective at predicting outcomes, it is important to note that such predictions are based on probabilities rather than certainties. Therefore, although these models can provide valuable guidance, they cannot ensure specific individual results. They are intended to supplement the clinical judgement of healthcare professionals, not substitute for it.

(ii) Interactive Web System Development

The website is designed to simplify the instructions and operability of the website. The simple layout and design of the "My Heart ACS Air" are user friendly and versatile. The system includes functionalities like prediction calculators and interactive databases. All menus and control buttons were included and labelled clearly for user navigation. Instructions are simple to understand, and descriptions and explanations are provided for users who are unfamiliar with how to use the website. A disclaimer notice is also included on each page of the website.

Two separated modules of admin and user are available. It increases the ease of the management process since the administrator can manage the registered users via “Users” page on the website. Therefore, all user identifications are verified and without the occurring of illegal user that improperly used the system and ensures the privacy and confidentiality of user data.

The system is capable to store and insert the patient’s data into the database. It allows user to manage patient information, functions such edit, delete, and update patients’ record. This is convenient for the users to make a follow-up on the condition of the patient after a certain time interval.

The creation of the "My Heart ACS Air" web application allows for more accessible, user interaction with the results of the research, making it easier for both medical professionals and the public to understand the correlation between air pollution and ACS. This can lead to increased awareness, informed decision-making, and proactive measures by the public regarding their health.

In addition, the web system is accessible via mobile phone, allowing users to quickly navigate to the system; the print screen of the mobile version interface is included in Appendix J.

(iii) Comprehensive Dataset

This study makes use of extensive dataset from the National Cardiovascular Database (NCVD) and air quality data from Department of Environment (DOE) Malaysia that covers a period of 12 years (2006–2017). The National Heart Association of Malaysia (NHAM) and the DOE Malaysia were responsible for its curation and management. Moreover, the NCVD-ACS patient information has been anonymized to protect their privacy and confidentiality.

The extensive duration of data collection allows the ML models to be trained on a wide range of various instances, which enhances the accuracy and reliability of the predictions.

(iv) Enhanced Visualization Tool

The “My Heart ACS Air” system consists of spatial visualization tool. Users are able choose the hospital location and input corresponding air quality data. In response, the system generates an interactive Google Map display that illustrates the predicted ACS hospitalization and mortality cases at the selected hospital. In addition, the "Hospital (Multiple) Geo-Location Prediction" page, allows for simultaneous visualization of multiple hospital locations along with their predicted outcomes based on the user input.

The system also aids in identifying high-risk patients, categorizing any patient with a mortality risk exceeding 50% as a "high-risk patient." This level of data presentation, accessible and easy to understand, can facilitate a deeper comprehension of the study's implications among medical professionals and government policy makers. This combination of specific geolocation-based prediction and risk categorization underscores the practical utility and user-friendly nature of this visualization tool.

(v) Resource Allocation Guidance and Public Health Awareness

The outcomes of this study hold profound implications for public health, environmental protection efforts, and resource allocation in healthcare. By underlining the impact of air pollution on ACS incidence, the study could potentially enlighten public awareness and guide policymaking in both public health and environmental domains. It emphasises the need for integrated health management that considers individual and environmental factors.

Based on the validated ML models, the study's predictions can help allocate healthcare resources to regions with a high concentration of ACS high-risk patients. This is significant because Malaysia has few well-equipped cardiac care facilities. Thus, the findings can improve the delivery of healthcare by guiding resource allocation., possibly improved ACS management.

Though the research is primarily focused on the Malaysian context, its implications may also resonate with countries at similar stages of their evolution in cardiovascular healthcare delivery. The demonstrated impact of this study, thus, extends beyond immediate clinical applications and can potentially influence broader strategies for air pollution control and cardiovascular disease management, leading to improved public health outcomes not only in Malaysia but also in comparable contexts globally.

5.5 Limitations of the Current Study

While the current research offers significant insights, it is also necessary to acknowledge the limitations inherent in our study that need to be addressed in future studies, the limitations are as follows:

(i) Dependence on Manual Data Input and Limited Real-Time Data

The primary limitation of the study is that the system depends on the manual input of air quality data by users. This requirement not only introduces potential inaccuracies and inconsistencies in the data but also adds a layer of complexity and redundancy for the users. Although a link to the air quality webpage for ease of data access is provided, the process of manual data entry is still required. Due to security restrictions that prevent the automatic acquisition of data readings from third-party sources, the data entry process is more cumbersome and less user-friendly than desired.

Additionally, the inability to automatically source real-time data from the Department of Environment limits the system's effectiveness in providing immediate and up-to-date risk assessments.

(ii) Limited Scope of Environmental Factors

The limited range of environmental factors incorporated into the model is one of the study’s limitations. The lack of Particulate Matter 2.5 (PM2.5) and Carbon Monoxide (CO) data in our dataset limits the scope of environmental considerations in our analysis. This is crucial that PM2.5 and CO are prominent pollutants that have been evidenced to impact the onset of ACS (Liu, et al., 2022; Qiu, et al., 2020; Zhao, et al., 2016; Meng, et al., 2016; Qorbani, et al., 2012).

The unavailability of PM2.5 and CO data is due to the air quality monitoring station unable to capture these particular reading during the research period (2006 – 2017) (Department of Environment Malaysia, 2021). Therefore, the lack of these key pollutants in our environmental considerations potentially restricts the full scope of our ACS risk assessment model, underscoring a crucial area for future research expansion.

5.6 Future Study

The methodology and findings can be utilized in future research to evaluate the impact of other environmental factors on various health conditions, potentially contributing to the broader field of environmental health research. Additionally, the findings can be used to inform public health policies related to air pollution control and cardiovascular disease management, potentially leading to improved public health outcomes in Malaysia, perhaps in Southeast Asia region as well.

Based on the results and implications from this study, several suggestions and enhancements can be worked on the future work in the project to improve the efficiency and the performance of the model and system developed. The recommended enhancements are described in the following:

(i) Develop Data Automation

Future research should prioritise the development of data automation, particularly by integrating an Application Programming Interface (API) or an automated system to retrieve real-time air quality data from the DOE Malaysia. Thus, enhance user convenience instead of manual data input and increase the accuracy of predictions by providing real-time environment information for the ML models. The process would require planning and collaboration with the relevant department to ensure consistent information extraction, making the system more user-friendly and reliable. Such improvements might substantially enhance the system's real-time prediction, improving our model's comprehensiveness and accuracy in predicting ACS onset.

(ii) Further Validation of Predictive Models

It is recommended in future research should conduct further validation and optimization of the ML models used in this study, testing on real-word current dataset. Comparison across diverse geographical and demographic contexts such as rural and urban hospitals in relation to local air quality readings would be able to provide new findings as well.

Furthermore, apart from validation using TIMI risk score, it is encouraged to validate the models against other reputable risk scores, such as GRACE risk score. The effort in continuous validation of risk score will enhance the models' predictive accuracy and reliability.

(iii) Regular Upgrades to Web System and ML Models

The web system and ML models should be subject to regular updates and upgrades, this will ensure that the tools remain accurate, relevant, and user-centric enhancing its overall usability and impact.

The model could be update and retrain with the most recent datasets, which keeps the validity and efficacy of the models in will with the varying real-world conditions and relevant to current public health conditions.

Besides, in the future can further work on the expansion of dataset parameter. The inclusion of additional environmental and meteorological variables into the model in subsequent studies is encouraged. Incorporating measures such as PM2.5, CO, temperature, and humidity would provide a broader perspective on the understanding of the diverse environmental factors influencing the onset of ACS, further enhancing healthcare precision.

For web system improvements could include practical suggestion following high-risk prediction. Areas with high forecasted admission rates or severe cardiac risk, the system could suggest strategic next steps such as directing resources to the nearest cardiac laboratories or suggest hospitalization monitoring for high-risk patients, these upgrades will eventually improve the overall utility and contribution.

(iv) Collaboration with Hospitals and Policymakers

Future research should suggest and facilitate the application of research findings in hospitals and among policymakers. By putting these findings into practice, ACS prevention, treatment, and management strategies may be improved on a larger scale.

CHAPTER 6:CONCLUSION

This study has successfully met its primary goals, providing insights on the relationship between Acute Coronary Syndrome (ACS) hospitalization, mortality rates, and air pollution in Malaysia. It has demonstrated the efficacy of Machine Learning (ML) models in predicting in-hospital mortality risk among ACS patients, using ACS and air quality data. These developments have been cohesively integrated into a web system, offering a visual representation of the potential health effects of air pollution.

By integrating environmental factors to healthcare predictive models and creating an interactive web platform, these models can predict air pollution-related hospitalisation and mortality risks, revolutionising healthcare delivery. The visualization tool equips healthcare providers and policymakers with crucial data, enabling a deeper understanding of the impact of air pollution on ACS hospitalization and mortality.

While this study is a significant step forward, its scope was limited by the timeframe of the input data and was limited to patients with ACS. The development and deployment of a system within hospital settings that allows direct data collection from healthcare professional users is an essential component of this future application. This system will allow for the continuous collection of patient data, significantly expanding our dataset for the model's ongoing refinement and practical application. Considerable additional research should be conducted on validating and updating the ML models with the most recent readings, allowing the ML models and web system to remain relevant and accurately reflect current trends. This study contributes to the larger goal of understanding and mitigating the health effects of air pollution by expanding the utility and applicability of these models.

REFERENCES

Abdullah, S., Napi, N., Ahmed, A., Mansor, W., Mansor, A., Ismail, M., . . . Ramly, Z. (2020). Development of multiple linear regression for particulate matter (PM10) forecasting during episodic transboundary haze event in Malaysia. *Atmosphere, 11(3)*, Article#289.

Afroz, R., Hassan, M. N., & Ibrahim, N. A. (2003). Review of air pollution and health impacts in Malaysia. *Environmental research, 92(2)*, 71-77.

Aghamohammadi, N., & Isahak, M. (2018). Climate change and air pollution in Malaysia. Climate Change and Air Pollution: The Impact on Human Health in Developed and Developing Countries. *Springer*, 241-254.

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. A., Zambahari, R., Ismail, O., Sinnadurai, J., Rosman, A., Piaw, C. S., . . . 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.

Ahmad., W. W. (2022). *Annual Report of the NCVD-ACS Registry, 2018–2019.* Kuala Lumpur: National Heart Association of Malaysia (NHAM).

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.

Ahsan, M. M., & Siddique, Z. (2022). Machine learning-based heart disease diagnosis: A systematic literature review. *Artificial Intelligence in Medicine*, Article#102289.

Ajitesh, K. (2023, March 21). *K-Fold Cross Validation – Python Example*. Retrieved from Data Analytics: AI, Data, Data Science, Machine Learning, Blockchain, Digital: https://vitalflux.com/k-fold-cross-validation-python-example/

Akyuz, S., Yazici, S., Bozbeyoglu, E., Onuk, T., Yildirimturk, O., Karacimen, D., . . . Cagdas, M. (2016). Validity of the updated GRACE risk predictor (version 2.0) in patients with non-ST-elevation acute coronary syndrome. *Revista Portuguesa de Cardiologia, 35(1)*, 25-31.

Al'Aref, S., Anchouche, K., Singh, G., Slomka, P., Kolli, K., Kumar, A., . . . Berman, D. (2019). Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. *European heart journal, 40(24)*, 1975-1986.

Alaa, A., Bolton, T., Di Angelantonio, E., Rudd, J., & Van der Schaar, M. (2019). Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants. *PloS one, 14(5)*, Article#e0213653.

Alahmar, A., Mohammed, E., & Benlamri, R. (2018). Application of data mining techniques to predict the length of stay of hospitalized patients with diabetes. *2018 4th International Conference on Big Data Innovations and Applications (Innovate-Data)* (pp. 38-43). Barcelona: IEEE.

Alhassan, S. M., Ahmed, H. G., Almutlaq, B. A., Alanqari, A. A., Alshammari, R. K., & Alshammari, K. T. (2017). Risk factors associated with acute coronary syndrome in northern Saudi Arabia. *Search of a Perfect Outfit. J Cardiol Curr Res, 8(3)*, Article#00281.

Alshamrani, A., & Bahattab, A. (2015). A comparison between three SDLC models waterfall model, spiral model, and Incremental/Iterative model. *International Journal of Computer Science Issues (IJCSI), 12(1)*, Article#106.

Altexsoft. (2022, July 26). *Non-functional Requirements: Examples, Types, How to Approach*. Retrieved from altexsoft: https://www.altexsoft.com/blog/non-functional-requirements/

Alty, S., Millasseau, S., Chowienczyc, P., & Jakobsson, A. (2003). Cardiovascular disease prediction using support vector machines. *2003 46th Midwest Symposium on Circuits and Systems (Vol. 1)* (pp. 376-379). IEEE.

Ameer, S., Shah, M., Khan, A., S. H., Maple, C., Islam, S., & Asghar, M. (2019). Comparative analysis of machine learning techniques for predicting air quality in smart cities. *IEEE Access, 7*, 128325-128338.

Amsterdam, E., Wenger, N., Brindis, R., Casey Jr, D., Ganiats, T., Holmes Jr, D., . . . Levine, G. (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.

Androutsopoulos, I., Koutsias, J., Chandrinos, K. V., Paliouras, G., & Spyropoulos, C. D. (2000). An evaluation of naive bayesian anti-spam filtering. *arXiv preprint.*, cs/0006013.

Angraal, S., Mortazavi, B., Gupta, A., Khera, R., Ahmad, T., Desai, N., . . . Krumholz, H. (2020). Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction. *JACC: Heart Failure, 8(1)*, 12-21.

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., Cohen, M., Bernink, P., McCabe, C., Horacek, T., Papuchis, G., . . . 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.

Antman, E., Cohen, M., Bernink, P., McCabe, C., Horacek, T., Papuchis, G., . . . 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.

Aottiwerch, N., & Kokaew, U. (2017). Design computer-assisted learning in an online Augmented Reality environment based on Shneiderman's eight Golden Rules. *2017 14th International Joint Conference on Computer Science and Software Engineering (JCSSE)* (pp. 1-5). IEEE.

Arjaria, S. K., Rathore, A. S., & Cherian, J. S. (2021). Kidney disease prediction using a machine learning approach: A comparative and comprehensive analysis. *Demystifying big data, machine learning, and deep learning for healthcare analytics* (pp. 307-333). Academic Press.

Arnett, D., Blumenthal, R., Albert, M., Buroker, A., Goldberger, Z., Hahn, E., . . . Michos, E. (2019). 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guideline. *Circulation, 140(11)*, e563-e595.

Ashenden, S. K. (2021). *The era of artificial intelligence, machine learning, and data science in the pharmaceutical industry.* United Kingdom: Academic Press.

Asia Pacific Cohort Studies, C. (2005). A comparison of the associations between risk factors and cardiovascular disease in Asia and Australasia. *European Journal of Preventive Cardiology, 12(5)*, 484-491.

Asif, M., Nishat, M., Faisal, F., Dip, R., Udoy, M., Shikder, M., & Ahsan, R. (2021). Performance Evaluation and Comparative Analysis of Different Machine Learning Algorithms in Predicting Cardiovascular Disease. . *Engineering Letters, 29(2)*, 1-7 .

Authors/Task Force Members, Steg, P., James, S., Atar, D., Badano, L., Lundqvist, C., . . . Fernandez-Aviles, F. (2012). ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: The Task Force on the management of ST-segment elevation acute myocardial infarction of the European Society of Cardiology (ESC). *uropean heart journal, 33(20)*, 2569-2619.

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.

Aziz, F., Malek, S., Ibrahim, K. K., & Kasim, S. (2019). A Novel Local Machine Learning Algorithm to Predict Death in ACS Patients. *International Journal of Cardiology., 297*, Article#18.

Aziz, F., Malek, S., Ibrahim, K., Raja Shariff, R., Wan Ahmad, W., Ali, R., . . . Kasim, S. (2021). Short-and long-term mortality prediction after an acute ST-elevation myocardial infarction (STEMI) in Asians: A machine learning approach. *PloS one, 16(8)*, Article#e0254894.

Bae, H. J. (2014). Effects of Short-term Exposure to PM 10 and PM 2.5 on Mortality in Seoul. *Journal of Environmental Health Sciences, 40(5)*, 346-354.

Bai, B., Tan, Y., Donchyts, G., Haag, A., Xu, B., Chen, G., & Weerts, A. H. (2023). Naive Bayes classification-based surface water gap-filling from partially contaminated optical remote sensing image. *Journal of Hydrology, 616*, Article#128791.

Bañeras, J., Ferreira-González, I., Marsal, J., Barrabés, J., Ribera, A., Lidón, R., . . . García-Dorado, D. (2018). Short-term exposure to air pollutants increases the risk of ST elevation myocardial infarction and of infarct-related ventricular arrhythmias and mortality. *International journal of cardiology, 250*, 35-42.

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an interpretive scale to the System Usability Scale. *Journal of Usability Studies, 4(3)*, 1-19.

Basra, S., Virani, S., Paniagua, D., Kar, B., & Jneid, H. (2016). Acute Coronary Syndromes: Unstable Angina and Non–ST Elevation Myocardial Infarction. *Heart Failure Clinics, 12(1)*, 31-48.

Batista, G. E., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD explorations newsletter, 6(1)*, 20-29.

Battineni, G., Chintalapudi, N., & Amenta, F. (2019). Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM). *Informatics in Medicine Unlocked, 16*, Article#100200.

Bawamia, B., Mehran, R., Qiu, W., & Kunadian, V. (2013). Risk scores in acute coronary syndrome and percutaneous coronary intervention: a review. *American heart journal, 165(4)*, 441-450.

Bayes, T. (1968). Naive bayes classifier. *Article Sources and Contributors*, 1-9.

Belyadi, H., & Haghighat, A. (2021). *Machine Learning Guide for Oil and Gas Using Python: A Step-by-Step Breakdown with Data, Algorithms, Codes, and Applications.* Gulf Professional Publishing.

Bennett, C. C., & Hauser, K. (2013). Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach. *Artificial intelligence in medicine, 57(1)*, 9-19.

Bergmann, S., Li, B., Pilot, E., Chen, R., Wang, B., & Yang, J. (2020). Effect modification of the short-term effects of air pollution on morbidity by season: A systematic review and meta-analysis. *Science of The Total Environment, 716*, Article#136985.

Berrar, D. (2018). Bayes' theorem and naive Bayes classifier. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics, 403*, Article#412.

Beswick, A. D., 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). *Royal College of General Practitioners (UK), London*, Article#21834196 .

Beunza, J., Puertas, E., García-Ovejero, E., Villalba, G., Condes, E., Koleva, G., . . . Landecho, M. (2019). Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease). *Journal of biomedical informatics, 97*, Article#103257.

Beverland, I., Cohen, G., Heal, M., Carder, M., Yap, C., Robertson, C., . . . Agius, R. (2012). A comparison of short-term and long-term air pollution exposure associations with mortality in two cohorts in Scotland. *Environmental health perspectives, 120(9)*, 1280-1285.

Bewick, V., Cheek, L., & Ball, J. (2005). Statistics review 14: Logistic regression. *Critical care, 9(1)*, Article#112.

Bhat, A. (2018). *What is System Usability Scale?* Retrieved from questionPro: questionpro.com/blog/system-usability-scale/

Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., & Lloyd, S. (2017). Quantum machine learning. *Nature, 549(7671)*, 195-202.

Biecek, P., & Burzykowski, T. (2021). *Explanatory model analysis: explore, explain, and examine predictive models.* CRC Press.

Biggeri, A., Bellini, P., & Terracini, B. (2004). Meta-analysis of the Italian studies on short-term effects of air pollution--MISA 1996-2002. *Epidemiologia e prevenzione, 28(4-5 Suppl)*, 4-100.

Boogaard, H., Walker, K., & Cohen, A. J. (2019). Air pollution: the emergence of a major global health risk factor. *International health, 11(6)*, 417-421.

Boukerche, F., Hammou, L., & Laredj, N. (2023). Performance of GRACE risk score for predicting 5-year cardiovascular mortality in NSTE-ACS patients. *Archives of Cardiovascular Diseases Supplements, 15(1)*, 15-16.

Bourdrel, T., Bind, M., Béjot, Y., & Morel, O. A. (2017). Cardiovascular effects of air pollution. *Archives of cardiovascular diseases, 110(11)*, 634-642.

Bouzas Cruz, N., Cordero, A., Alvarez-Alvarez, B., Bertomeu-Gonzalez, V., Gonzalez-Ferrero, T., Zuazola, P., . . . Diaz-Louzao, C. (2021). The value of GRACE risk score for predicting mortality in heart failure patients admitted with non-ST elevation acute coronary syndrome. *European Heart Journal, 42(Supplement_1)*, ehab724-0808.

Bradstreet, J. W. (1995). Hazardous air pollutants: Assessment, liabilities, and regulatory compliance. *Noyes Publications*, 1-7.

Braun, H., Patterson, D., Molloy, A., & Davies, K. (2020). Predicting mortality risk in patients with coronavirus or influenza using artificial intelli-gence. *infection, 7*, Article#8.

Breezometer. (2023). *Delivering the World's Most Accurate Air Quality Data*. Retrieved from BreezoMeter: https://www.breezometer.com/accurate-realtime-air-quality-data

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

Breiman, L. (2001). Machine learning. *Springer (Vol. 45)*, 5-32.

Breiman, L. (2001). Random forests. *Machine learning, 45(1)*, 5-32.

Briggs, D., Collins, S., Elliott, P., Fischer, P., Kingham, S., Lebret, E., . . . Van Der Veen, A. (1997). Mapping urban air pollution using GIS: a regression-based approach. *International Journal of Geographical Information Science, 11(7)*, 699-718.

Brook, R., Franklin, B., Cascio, W., Hong, Y., Howard, G., Lipsett, M., . . . Tager, I. (2004). Air pollution and cardiovascular disease: a statement for healthcare professionals from the Expert Panel on Population and Prevention Science of the American Heart Association. *Circulation, 109(21)*, 2655-2671.

Brook, R., Rajagopalan, S., Pope III, C., Brook, J., Bhatnagar, A., Diez-Roux, A., . . . Peters, A. (2010). Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association. *Circulation 121, no. 21*, 2331-2378.

Brooke, J. (1986). System usability scale (SUS): a quick-and-dirty method of system evaluation user information. *Reading, UK: Digital equipment co ltd, 43*, 1-7.

Brooke, J. (1996). SUS: A quick and dirty usability scale. *Usability evaluation in industry*, 189-194.

Brown, H. (2005). A web-based system to monitor and predict healthcare activity. *Health Informatics Journal, 11(1)*, 63-79.

Bueno, H., & Fernández-Avilés, F. (2012). Use of risk scores in acute coronary syndromes. *Heart, 98(2)*, 162-168.

Buja, M. L., & Butany, J. (2022). *Cardiovascular pathology.* Academic Press.

Bulluck, H., Zheng, H., Chan, M., Foin, N., Foo, D., Lee, C., . . . Tong, K. (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*, 1-14.

Butland, B., Atkinson, R., Milojevic, A., Heal, M., Doherty, R., Armstrong, B., . . . Wilkinson, P. (2016). Myocardial infarction, ST-elevation and non-ST-elevation myocardial infarction and modelled daily pollution concentrations: a case-crossover analysis of MINAP data. *Open Heart, 3(2)*, Article#e000429.

Castro-Dominguez, Y., Dharmarajan, K., & McNamara, R. L. (2018). Predicting death after acute myocardial infarction. *Trends in Cardiovascular Medicine, 28(2)*, 102-109.

Chan, M., Shah, B., Gao, F., Sim, L., Chua, T., Tan, H., . . . Surrun, S. (2011). Recalibration of the Global Registry of Acute Coronary Events risk score in a multiethnic Asian population. *American heart journal, 162(2)*, Article#291-299.

Chatterjee, A., Gerdes, M. W., & Martinez, S. G. (2020). Identification of risk factors associated with obesity and overweight—a machine learning overview. *Sensors, 20(9)*, Article#2734.

Cheema, F., Cheema, H., & Akram, Z. (2020). Identification of risk factors of acute coronary syndrome in young patients between 18-40 years of age at a teaching hospital. *Pakistan Journal of Medical Sciences, 36(4)*, Article#821.

Chen, J., & Hoek, G. (2020). Long-term exposure to PM and all-cause and cause-specific mortality: a systematic review and meta-analysis. *Environment international, 143*, Article#105974.

Chen, J., Xing, Y., Xi, G., Chen, J., Yi, J., Zhao, D., & Wang, J. (2007). A comparison of four data mining models: Bayes, neural network, SVM and decision trees in identifying syndromes in coronary heart disease. *International Symposium on Neural Networks* (pp. 1274-1279). Berlin: Springer.

Chen, R., Jiang, Y., Hu, J., Chen, H., Li, H., Meng, X., . . . Fang, W. (2022). Hourly air pollutants and acute coronary syndrome onset in 1.29 million patients. *Circulation, 145(24)*, 1749-1760.

Chen, R., Yin, P., Meng, X., Wang, L., Liu, C., Niu, Y., . . . You, J. (2018). Associations between ambient nitrogen dioxide and daily cause-specific mortality: evidence from 272 Chinese cities. *Epidemiology, 29(4)*, 482-489.

Chen, S., Webb, G. I., Liu, L., & Ma, X. (2020). A novel selective naïve Bayes algorithm. *Knowledge-Based Systems, 192*, Article#105361.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, (pp. 785-794).

Chen, X., Wu, H., Li, L., Zhao, X., Zhang, C., & Wang, W. E. (2022). The prognostic utility of GRACE risk score in predictive adverse cardiovascular outcomes in patients with NSTEMI and multivessel disease. *BMC Cardiovascular Disorders, 22(1)*, 1-8.

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)*, Article#4.

Cheng, F., Wu, K., Hung, S., Lee, K., Lee, C., Liu, K., & Hsu, P. (2020). Association between ambient air pollution and out-of-hospital cardiac arrest: are there potentially susceptible groups? *Journal of Exposure Science & Environmental Epidemiology, 30(4)*, 641-649.

Chew, D. P., Scott, I. A., Cullen, L., French, J. K., Briffa, T. G., Tideman, P. A., . . . Group, N. A. (2016). National Heart Foundation of Australia & Cardiac Society of Australia and New Zealand: Australian Clinical Guidelines for the Management of Acute Coronary Syndromes 2016. *Heart, lung & circulation, 25(9)*, 895–951.

Chong, K., & Shah, N. (2022). Comparison of Naive Bayes and SVM Classification in Grid-Search Hyperparameter Tuned and Non-Hyperparameter Tuned Healthcare Stock Market Sentiment Analysis. *International Journal of Advanced Computer Science and Applications, 13(12).* (pp. 90-94). Science and Information (SAI) Organization Limited.

Chuang, K. J., Yan, Y. H., Chiu, S. Y., & Cheng, T. J. (2011). Long-term air pollution exposure and risk factors for cardiovascular diseases among the elderly in Taiwan. *Occupational and environmental medicine, 68(1)*, 64-68.

Collet, J., Thiele, H., Barbato, E., Barthélémy, O., Bauersachs, J., Bhatt, D., . . . Gale, C. (2021). 2020 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: The Task Force for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation. *European heart journal, 42(14)* , 1289-1367.

Conrad, E., Misenar, S., & Feldman, J. (2016). *Eleventh Hour CISSP®: Study Guide.* Syngress.

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.

Cornell, S., Doust, J., Morgan, M., Greaves, K., Hawkes, A., de Wet, C., . . . Bonner, C. (2023). Implementing patient decision aids into general practice clinical decision support systems: Feasibility study in cardiovascular disease prevention. *PEC Innovation, 2*, Article#100140.

Correia, L. C., Garcia, G., Kalil, F., Ferreira, F., Carvalhal, M., Oliveira, R., . . . 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*, 273-297.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological), 20(2)*, 215-232.

Cuadros, D. F., Xiao, Y., Mukandavire, Z., Correa-Agudelo, E., Hernández, A., Kim, H., & MacKinnon, N. J. (2020). Spatiotemporal transmission dynamics of the COVID-19 pandemic and its impact on critical healthcare capacity. *Health & place, 64*, Article#102404.

Daga, L. C., Kaul, U., & Mansoor, A. (2011). Approach to STEMI and NSTEMI. *J Assoc Physicians India, 59(Suppl 12)*, 19-25.

Dagliati, A. M., Chiovato, L., & Bellazzi, R. (2018). Machine learning methods to predict diabetes complications. *Journal of diabetes science and technology, 12(2)*, 295-302.

Damen, J., Hooft, L., Schuit, E., Debray, T., Collins, G., Tzoulaki, I., . . . Schlüssel, M. (2016). Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ*, Article#353.

Dastoorpoor, M., Sekhavatpour, Z., Masoumi, K., Mohammadi, M., Aghababaeian, H., Khanjani, N., . . . Vahedian, M. (2019). Air pollution and hospital admissions for cardiovascular diseases in Ahvaz, Iran. *Science of the total environment, 652*, 1318-1330.

De Marco, A., Proietti, C., Anav, A., Ciancarella, L., D'Elia, I., Fares, S., . . . Marchetto, A. (2019). Impacts of air pollution on human and ecosystem health, and implications for the National Emission Ceilings Directive: Insights from Italy. *Environment International, 125*, 320-333.

de Vries, M. J., Land-Zandstra, A. M., & Smeets, I. (2019). Citizen scientists' preferences for communication of scientific output: a literature review. *Citizen Science: Theory and Practice, 4(1)*, Article#2.

Demir, F. (2022). Deep autoencoder-based automated brain tumor detection from MRI data. *Artificial Intelligence-Based Brain-Computer Interface* (pp. 317-351). Academic Press.

Deng, S., Wang, Z., Zhang, Y., Xin, Y., Zeng, C., & Hu, X. (2022). Association of fasting blood glucose to high-density lipoprotein cholesterol ratio with short-term outcomes in patients with acute coronary syndrome. *Lipids in Health and Disease, 21(1)*, 1-9.

Department of Environment Malaysia, D. (2021, October). *Air Quality Monitoring Station in Malaysia.* Retrieved from Official Portal Department of Environment Ministry if Environment and Water: https://www.doe.gov.my/en/air-quality-monitoring-station-in-malaysia/

Di, Q., Dai, L., Wang, Y., Zanobetti, A., Choirat, C., Schwartz, J. D., & Dominici, F. (2017). Association of short-term exposure to air pollution with mortality in older adults. *Jama, 318(24)*, 2446-2456.

Díaz-Chirón, L., Negral, L., Megido, L., Suárez-Peña, B., Domínguez-Rodríguez, A., Rodríguez, S., . . . Avanzas, P. (2021). Relationship between exposure to sulphur dioxide air pollution, white cell inflammatory biomarkers and enzymatic infarct size in patients with ST-segment elevation acute coronary syndromes. . *European Cardiology Review, 16*, 1-10.

Díaz-Uriarte, R., & De Andres, S. A. (2006). Gene selection and classification of microarray data using random forest. *BMC bioinformatics, 7(1)*, Article#3.

Dickerson, P. (2012). AIRNOW. *2012 Socio-economic Benefits Workshop: Defining, measuring, and Communicating the Socio-economic Benefits of Geospatial Information* (pp. 1-6). IEEE.

Dietterich, T. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation, 10(7)*, 1895-1923.

Dikshit, A., & Pradhan, B. (2021). Interpretable and explainable AI (XAI) model for spatial drought prediction. *Science of the Total Environment, 801*, Article#149797.

Dinh, A., Miertschin, S., Young, A., & Mohanty, S. (2019). A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. *BMC medical informatics and decision making, 19(1)*, 1-15.

Dockery, D. W., & Pope, C. A. (1994). Acute respiratory effects of particulate air pollution. *Annual review of public health, 15(1)*, 107-132.

Domingues, I., Amorim, J. P., Abreu, P. H., Duarte, H., & Santos, J. (2018). Evaluation of oversampling data balancing techniques in the context of ordinal classification. *2018 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.

Dominguez-Rodriguez, A., Abreu-Gonzalez, P., Rodríguez, S., Avanzas, P., & Juarez-Prera, R. A. (2017). Short-term effects of air pollution, markers of endothelial activation, and coagulation to predict major adverse cardiovascular events in patients with acute coronary syndrome: insights from AIRACOS study. *Biomarkers, 22(5)*, 389-393.

DOSM, & Portal, D. o. (2021, November 16). *Statistics on Causes of Death, Malaysia, 2021*. Retrieved from Department of Statistics Malaysia Officail Portal: The Source of Malaysia's Official Statistics: https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=401&bul_id=R3VrRUhwSXZDN2k4SGN6akRhTStwQT09&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT09

DOSM, D. o. (2022). *Statistics on Causes of Death, Malaysia, 2022.* Putrajaya: Department of Statistics Malaysia.

Du, X., Huang, T., & Wang, S. (2023). Predicting mortality in patients with heart failure based on machine learning approaches. *Second International Conference on Biological Engineering and Medical Science (ICBioMed 2022) (Vol. 12611)* (pp. 1217-1227). SPIE.

Du, Y., Xu, X., Chu, M., Guo, Y., & Wang, J. (2016). Air particulate matter and cardiovascular disease: the epidemiological, biomedical and clinical evidence. *Journal of thoracic disease, 8(1)*, Article#E8.

Dutta, P., Paul, S., Cengiz, K., Anand, R., & Majumder, M. (2023). A predictive method for emotional sentiment analysis by machine learning from electroencephalography of brainwave data. *Implementation of Smart Healthcare Systems using AI, IoT, and Blockchain* (pp. 109-130). Academic Press.

Earl, M. J. (1978). Prototype systems for accounting, information and control. . *Accounting, Organizations and Society, 3(2)*, 161-170.

Edgar, T., & Manz, D. (2017). *Research methods for cyber security.* Syngress.

Eggers, K. M., Baron, T., Hjort, M., Nordenskjöld, A. M., Tornvall, P., & Lindahl, B. (2021). GRACE 2.0 Score for Risk Prediction in Myocardial Infarction With Nonobstructive

Coronary Arteries. *Journal of the American Heart Association, 10(17)*, Article#e021374.

Ekanayake, I. U., Meddage, D. P., & Rathnayake, U. (2022). A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP). *Case Studies in Construction Materials, 16*, Article#e01059.

El Bouchefry, K., & de Souza, R. S. (2020). Learning in big data: Introduction to machine learning. In Š. Petr, & A. Fathalrahman, *Knowledge discovery in big data from astronomy and earth observation* (pp. 225-249). Elsevier.

Elbarouni, B., Goodman, S., Yan, R., Welsh, R., Kornder, J., DeYoung, J., . . . Tan, M. (2009). Validation of the Global Registry of Acute Coronary Event (GRACE) risk score for in-hospital mortality in patients with acute coronary syndrome in Canada. *American heart journal, 158(3)*, 392-399.

Elhadd, T., Mall, R., Bashir, M., Palotti, J., Fernandez-Luque, L., Farooq, F., . . . PROFAST. (2020). Artificial Intelligence (AI) based machine learning models predict glucose variability and hypoglycaemia risk in patients with type 2 diabetes on a multiple drug regimen who fast during ramadan (The PROFAST–IT Ramadan study). *Diabetes Research and Clinical Practice, 169*, Article#108388.

Emakhu, J., Monplaisir, L., Aguwa, C., Arslanturk, S., Masoud, S., Nassereddine, H., . . . Miller, J. (2022). Acute coronary syndrome prediction in emergency care: A machine learning approach. *Computer methods and programs in biomedicine, 225*, Article#107080.

Esteban, M., Montero, S., Sánchez, J., Hernández, H., Pérez, J., Afonso, J., . . . de León, A. (2014). Acute coronary syndrome in the young: clinical characteristics, risk factors and prognosis. *The open cardiovascular medicine journal, 8*, Article#61.

Fabris, F., De Magalhães, J. P., & Freitas, A. A. (2017). A review of supervised machine learning applied to ageing research. *Biogerontology, 18(2)*, 171-188.

Faizal, A. S., 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*, Article#106190.

Fawagreh, K., Gaber, M. M., & Elyan, E. (2015). On extreme pruning of random forest ensembles for real-time predictive applications. *arXiv preprint*, Article#1503.04996.

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., . . . 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.

Feigin, V. L., Krishnamurthi, R., Merkin, A., Nair, B., Kravchenko, M., & Jalili-Moghaddam, S. (2022). Digital solutions for primary stroke and cardiovascular disease prevention: A mass individual and public health approach. *The Lancet Regional Health–Western Pacific 29* , 1-5.

Feng, S., Gao, D., Liao, F., Zhou, F., & Wang, X. (2016). The health effects of ambient PM2. 5 and potential mechanisms. *Ecotoxicology and environmental safety, 128*, 67-74.

Fierro, M. (2000). Particulate matter. *Air Info Now*, 1-11.

Fisher, E., & Korber, R. (2014). *BreezoMeter: Delivering global environmental information with Google Cloud*. Retrieved from Google Cloud: https://cloud.google.com/customers/breezometer

Fisher, S. B., Cropper, M., Kumar, P., Binagwaho, A., Koudenoukpo, J., Park, Y., . . . Landrigan, P. (2021). Air pollution and development in Africa: impacts on health, the economy, and human capital. *The Lancet Planetary Health, 5(10)*, e681-e688.

Fleury, V., Himsl, R., Joost, S., Nicastro, N., Bereau, M., Guessous, I., & Burkhard, P. R. (2021). Geospatial analysis of individual-based Parkinson's disease data supports a link with air pollution: A case-control study. *Parkinsonism & Related Disorders, 83*, 41-48.

Fox, K., Dabbous, O., Goldberg, R., Pieper, K., Eagle, K., Van de Werf, F., . . . Granger, C. (2006). Prediction of risk of death and myocardial infarction in the six months after presentation with acute coronary syndrome: prospective multinational observational study (GRACE). *bmj, 333(7578)*, Article#1091.

Fox, K., FitzGerald, G., Puymirat, E., Huang, W., Carruthers, K., Simon, T., . . . Anderson, F. (2014). Should patients with acute coronary disease be stratified for management according to their risk? Derivation, external validation and outcomes using the updated GRACE risk score. *BMJ open, 4(2)*, Article#e004425.

Franchini, M., & Mannucci, P. M. (2007). Short-term effects of air pollution on cardiovascular diseases: outcomes and mechanisms. *Journal of Thrombosis and Haemostasis, 5(11)*, 2169-2174.

Franchini, M., & Mannucci, P. M. (2012). Air pollution and cardiovascular disease. *Thrombosis research, 129(3)*, 230-234.

Frank, E., Trigg, L., Holmes, G., & Witten, I. H. (2000). Naive Bayes for regression. *Machine Learning, 41*, 5-25.

Franklin, B. A., Brook, R., & Pope III, C. A. (2015). Air pollution and cardiovascular disease. *Current problems in cardiology, 40(5)*, 207-238.

Fuster, V., & Kelly, B. B. (2010). *Epidemiology of cardiovascular disease. In Promoting cardiovascular health in the developing world: A critical challenge to achieve global health.* Washington (DC): National Academies Press (US).

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland, 15*, 246-263.

Gandhi, R. (2018, May 27). *Introduction to Machine Learning Algorithms: Linear Regression. Build your own model from scratch*. Retrieved from medium: https://towardsdatascience.com/introduction-to-machine-learning-algorithms-linear-regression-14c4e325882a

Garrett, G. (2023, April 16). *Quick list of useful R packages*. Retrieved from Postit Support: https://support.posit.co/hc/en-us/articles/201057987-Quick-list-of-useful-R-packages

Gestro, M., Condemi, V., Bardi, L., Tomaino, L., Roveda, E., Bruschetta, A., . . . Esposito, F. (2020). Short-term air pollution exposure is a risk factor for acute coronary syndromes in an urban area with low annual pollution rates: Results from a retrospective observational study (2011–2015). *Archives of Cardiovascular Diseases, 113(5)*, 308-320.

Ghaffar, A., Reddy, K. S., & Singhi, M. (2004). Burden of non-communicable diseases in South Asia. . *Bmj, 328(7443)*, 807-810.

Ghaffari, S. (2022). Non-ST-Elevation Acute Coronary Syndromes. In Practical Cardiology. *Elsevier*, 413-428.

Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation. *Scholarworks@UTEP*, Article#1209.

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*, 1-9.

Gillis, N., Arslanian-Engoren, C., & Struble, L. (2014). Acute coronary syndromes in older adults: a review of literature. *Journal of Emergency Nursing, 40(3)*, 270-275.

Giri, J., Raut, S., Rimal, B., Adhikari, R., Joshi, T. P., & Shah, G. (2023). Impact of air pollution on human health in different geographical locations of Nepal. *Environmental Research*, Article#115669.

Gjeka, R., Patel, K., Reddy, C., & Zetsche, N. (2021). Patient engagement with digital disease management and readmission rates: the case of congestive heart failure. *Health Informatics Journal, 27(3)*, Article#14604582211030959.

Goff Jr, D., Lloyd-Jones, D., Bennett, G., Coady, S., D’agostino, R., Gibbons, R., . . . Robinson, 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. *Circulation, 129(25_suppl_2)*, S49-S73.

Goldstein, B., Navar, A., & Carter, R. (2017). Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. *European heart journal, 38(23)*, 1805-1814.

Gómez-García, M. A., Pitchon, V., & Kiennemann, A. (2005). Pollution by nitrogen oxides: an approach to NOx abatement by using sorbing catalytic materials. *Environment international, 31(3)*, 445-467.

Gong, W., & Wang, S. (2009). Support vector machine for assistant clinical diagnosis of cardiac disease. *2009 WRI Global Congress on Intelligent Systems (Vol. 3)* (pp. 588-591). IEEE .

Gore, J., & Fox, A. A. (2023). *GRACE ACS Risk and Mortality Calculator*. Retrieved from MD+ Calc: https://www.mdcalc.com/calc/1099/grace-acs-risk-mortality-calculator

Goswami, S., Brady, J., Jordan, D., & Li, G. (2012). Intraoperative cardiac arrests in adults undergoing noncardiac surgery: incidence, risk factors, and survival outcome. *The Journal of the American Society of Anesthesiologists, 117(5)*, 1018-1026.

Goto, T., Camargo Jr, C. A., Faridi, M. K., Yun, B. J., & Hasegawa, K. (2018). Machine learning approaches for predicting disposition of asthma and COPD exacerbations in the ED. *The American journal of emergency medicine, 36(9)*, 1650-1654.

Goto, T., Camargo, C., Faridi, M., Freishtat, R., & Hasegawa, K. (2019). Machine learning–based prediction of clinical outcomes for children during emergency department triage. *JAMA network open, 2(1)*, e186937-e186937.

Graham, C. A., Tsay, S. X., Rotheray, K. R., & Rainer, T. H. (2013). Validation of the TIMI risk score in Chinese patients presenting to the emergency department with chest pain. *International journal of cardiology, 168(1)*, 597-598.

Grainger, S., Mao, F., & Buytaert, W. (2016). Environmental data visualisation for non-scientific contexts: Literature review and design framework. *Environmental Modelling & Software, 85*, 299-318.

Granger, C., Goldberg, R., Dabbous, O., Pieper, K., Eagle, K., Cannon, C., . . . Fox, K. (2003). Predictors of hospital mortality in the global registry of acute coronary events. *Archives of internal medicine 163, no. 19 ():* , 2345-2353.

Grier, R., Bangor, A., Kortum, P., & Peres, S. (2013). The system usability scale: Beyond standard usability testing. *Proceedings of the human factors and ergonomics society annual meeting (Vol. 57, No. 1)* (pp. 187-191). Los Angeles: SAGE Publications.

Gudivada, V. N. (2016). Cognitive computing: concepts, architectures, systems, and applications. *Handbook of statistics (Vol. 35)* (pp. 3-38). Elsevier.

Gupta, M. (2023, February 17). *Linear Regression in Machine learning*. Retrieved from GeeksforGeeks: https://www.geeksforgeeks.org/ml-linear-regression/

Gurjar, B., Jain, A., Sharma, A., Agarwal, A., Gupta, P., Nagpure, A., & Lelieveld, J. (2010). Human health risks in megacities due to air pollution. *Atmospheric Environment, 44(36)*, 4606-4613.

Hadjiev, D., Mineva, P., & Vukov, M. (2003). Multiple modifiable risk factors for first ischemic stroke: a population-based epidemiological study. *European Journal of Neurology, 10(5)*, 577-582.

Harangi, B., Antal, B., & Hajdu, A. (2012). Automatic exudate detection with improved Naïve-Bayes classifier. *2012 25th IEEE international symposium on computer-based medical systems (CBMS)* (pp. 1-4). IEEE.

Harrell, F. E. (2015). Binary logistic regression Regression modeling strategies. *Springer*, 219-274.

Harvard Chan Bioinformatics Core, H. (2023). *Packages and libraries*. Retrieved from HBC Training Github: https://hbctraining.github.io/Intro-to-R-flipped/lessons/04_introR_packages.html

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction (Vol. 2).* Springer.

Hertel, O., Johnson, M. S., & Goodsite, M. E. (2020). Air pollution sources, statistics, and health effects: Introduction. *Air Pollution Sources, Statistics and Health Effects*, 1-3.

Hess, E. P., Agarwal, D., Chandra, S., Murad, M. H., Erwin, P. J., Hollander, J. E., . . . Stiell, I. (2010). Diagnostic accuracy of the TIMI risk score in patients with chest pain in the emergency department: a meta-analysis. *Cmaj, 182(10)* , 1039-1044.

Hilbe, J. M. (2009). Logistic regression models. *CRC press*, 1.

Ho, T. (1995). Random decision forests. *3rd international conference on document analysis and recognition (Vol. 1)* (pp. 278-282). IEEE.

Hoek, G., Krishnan, R., Beelen, R., Peters, A., Ostro, B., Brunekreef, B., & Kaufman, J. (2013). Long-term air pollution exposure and cardio-respiratory mortality: a review. *Environmental health, 12(1)*, 1-16.

Holmberg, M. J., & Andersen, L. W. (2022). Adjustment for Baseline Characteristics in Randomized Clinical Trials. *JAMA, 328(21)*, 2155-2156.

Hongzong, S., Tao, W., Xiaojun, Y., Huanxiang, L., Zhide, H., Mancang, L., & BoTao, F. (2007). Support Vector Mechines Classification for Discriminating Coronary Heart Disease Patients from Non-coronary Heart Disease. *West Indian Medical Journal, 56(5)*, Article#451.

Hossen, M. A., Tazin, T., Khan, S., Alam, E., Sojib, H. A., Monirujjaman Khan, M., & Alsufyani, A. (2021). Supervised machine learning-based cardiovascular disease analysis and prediction. *Mathematical Problems in Engineering, 2021*, 1-10.

Hsieh, M., Lin, S., Lin, C., Hsieh, M., Hsu, W., Ju, S., . . . Kao, C. (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)*, Article#732.

Huang, C. H., Lin, H. C., Tsai, C. D., Huang, H. K., Lian, I. B., & Chang, C. C. (2017). The interaction effects of meteorological factors and air pollution on the development of acute coronary syndrome. *Scientific reports, 7(1)*, 1-10.

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.

Huang, J., Wei, X., Wang, Y., Jiang, M., Lin, Y., Su, Z., . . . Yu, D. (2021). Comparison of Prognostic Value Among 4 Risk Scores in Patients with Acute Coronary Syndrome: Findings from the Improving Care for Cardiovascular Disease in China-ACS (CCC-ACS) Project. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research, 27*, e928863-1.

Huang, W., FitzGerald, G., Goldberg, R., Gore, J., McManus, R., Awad, H., . . . Fox, K. (2016). Performance of the GRACE risk score 2.0 simplified algorithm for predicting 1-year death after hospitalization for an acute coronary syndrome in a contemporary multiracial cohort. *The American journal of cardiology, 118(8)*, 1105-1110.

Hughes, L. O., Raval, U., & Raftery, E. B. (1989). First myocardial infarctions in Asian and white men. *British medical journal, 298(6684)*, 1345-1350.

Huynh, Q., Blizzard, C., Marwick, T., & Negishi, K. (2018). Association of ambient particulate matter with heart failure incidence and all-cause readmissions in Tasmania: an observational study. *BMJ open, 8(5)*, Article#e021798.

Huynh, Q., Marwick, T., Venkataraman, P., Knibbs, L., Johnston, F., & Negishi, K. (2021). Long-term exposure to ambient air pollution is associated with coronary artery calcification among asymptomatic adults. *European Heart Journal-Cardiovascular Imaging, 22(8)*, 922-929.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting, 22(4)*, 679-688.

Imamovic, D., Babovic, E., & Bijedic, N. (2020). Prediction of mortality in patients with cardiovascular disease using data mining methods. *2020 19th International Symposium INFOTEH-JAHORINA (INFOTEH)* (pp. 1-4). IEEE.

Inmon, W. H., Linstedt, D., & Levins, M. (2019). *Data Architecture: A Primer for the Data Scientist: A Primer for the Data Scientist.* Academic Press.

IPH, I. f. (2020). *National Health and Morbidity Survey (NHMS) 2019: Non-communicable diseases, healthcare demand, and health literacy—Key Findings.* Malaysia: Institute for Public Health, National Institutes of Health (NIH).

Ismail, S., Khalil, M., Mohamad, M., & Azhar Shah, S. (2022). Systematic review and meta-analysis of prognostic models in Southeast Asian populations with acute myocardial infarction. *Frontiers in Cardiovascular Medicine*, Article#1850.

Jadhav, A., Mostafa, S. M., Elmannai, H., & Karim, F. K. (2022). An Empirical Assessment of Performance of Data Balancing Techniques in Classification Task. *Applied Sciences, 12(8)*, Article#3928.

Jain, D., & Singh, V. (2018). Feature selection and classification systems for chronic disease prediction: A review. *Egyptian Informatics Journal, 19(3)*, 179–189.

Jamthikar, A., Gupta, D., Mantella, L., Saba, L., Johri, A., & Suri, J. (2021). Ensemble Machine Learning and its Validation for Prediction of Coronary Artery Disease and Acute Coronary Syndrome using Focused Carotid Ultrasound. . *IEEE Transactions on Instrumentation and Measurement, 71*, 1-10.

Jamthikar, A., Gupta, D., Mantella, L., Saba, L., Laird, J., Johri, A., & Suri, J. (2021). Multiclass machine learning vs. conventional calculators for stroke/CVD risk assessment using carotid plaque predictors with coronary angiography scores as gold standard: A 500 participants study. . *The International Journal of Cardiovascular Imaging, 37(4)*, 1171-1187.

Jason, B. (2021, April 19). *A Gentle Introduction to Ensemble Learning Algorithms*. Retrieved from Machine Learning Mastery: https://machinelearningmastery.com/tour-of-ensemble-learning-algorithms/

JavaTpoint. (2011). *Machine learning Algorithms*. Retrieved from JavaTpoint: https://www.javatpoint.com/machine-learning

Jiang, Y., Huang, J., Li, G., Wang, W., Wang, K., Wang, J., . . . Guo, X. (2023). Ozone pollution and hospital admissions for cardiovascular events. *European Heart Journal, 44(18)*, 1622-1632.

Jiawei, H., Micheline, K., & Jian, P. (2012). *Data Mining: Concepts and Techniques.* -3rd. Morgan kaufmann.

John Lu, Z. Q. (2010). The elements of statistical learning: data mining, inference, and prediction. *Oxford University Press*, 693-694.

Jones, D. S. (2006). ASEAN and transboundary haze pollution in Southeast Asia. *Asia Europe Journal, 4(3)*, 431-446.

Joo, G., Song, Y., Im, H., & Park, J. (2020). Clinical implication of machine learning in predicting the occurrence of cardiovascular disease using big data (Nationwide Cohort Data in Korea). *IEEE Access, 8*, 157643-157653.

Joskow, P. L., Schmalensee, R., & Bailey, E. M. (1998). The market for sulfur dioxide emissions. *American Economic Review*, 669-685.

Juhan, N., Khalid, Z., Zubairi, Y., Zuhdi, A., & Ahmad, W. (2019). Risk factors of cardiovascular disease among st-elevation myocardial infarction male patients in Malaysia from 2006 to 2013. . *Jurnal Teknologi, 81(3)*, 145 - 149.

Kakadiaris, I. A., Vrigkas, M., Yen, A. A., Kuznetsova, T., Budoff, M., & Naghavi, M. (2018). Machine learning outperforms ACC/AHA CVD risk calculator in MESA. *Journal of the American Heart Association, 7(22)*, Article#e009476.

Kalcheva, N., Todorova, M., & Marinova, G. (2020). Naive Bayes Classifier, Decision Tree and AdaBoost Ensemble Algorithm–Advantages and Disadvantages. *Proceedings of the 6th ERAZ Conference Proceedings (part of ERAZ conference collection), Online* (pp. 153-157). ERAZ conference.

Kan, H., Chen, R., & Tong, S. (2012). Ambient air pollution, climate change, and population health in China. *Environment international, 42*, 10-19.

Kao, Y. T., Hsieh, Y. C., Hsu, C. Y., Huang, C. Y., Hsieh, M. H., Lin, Y. K., & Yeh, J. S. (2020). Comparison of the TIMI, GRACE, PAMI and CADILLAC risk scores for prediction of long-term cardiovascular outcomes in Taiwanese diabetic patients with ST-segment elevation myocardial infarction: From the registry of the Taiwan Society of Cardiology. *PLoS One, 15(2)*, Article#e0229186.

Karageorgou, D., Micha, R., & Zampelas, A. (2015). Mediterranean Diet and Cardiovascular Disease: An Overview of Recent Evidence. *The Mediterranean Diet*, 91-104.

Kasim, S. S., Malek, S., Ibrahim, K. S., Sureskumar, D., Aziz, M. F., Ibrahim, N., & Song, C. (2022a). Mortality prediction of elderly Asian patients with acute coronary syndrome (ACS) using interpretable machine learning algorithm. *International Journal of Cardiology 369*, 9-10.

Kasim, S., Malek, S., Cheen, S., Safiruz, M., Ahmad, W., Ibrahim, K., . . . Ibrahim, N. (2022b). In-hospital risk stratification algorithm of Asian elderly patients. *Scientific Reports, 12(1)*, 1-17.

Kasim, S., Malek, S., Ibrahim, K., Amir, P., & Aziz, M. (2021). 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.

Kasim, S., Malek, S., Song, C., Wan Ahmad, W., Fong, A., Ibrahim, K., . . . Ibrahim, N. (2022). In-hospital mortality risk stratification of Asian ACS patients with artificial intelligence algorithm. *PloS one, 17(12)*, Article#e0278944.

Kasim, S., Rudin, P., Malek, S., Ibrahim, K., Ahmad, W., Fong, A., . . . Ibrahim, N. (2023). In-Hospital Mortality Prediction using Machine Learning and Stacked Ensemble Learning of Asian Women with ST-Elevation Myocardial Infarction (STEMI). *Research Square*.

Kasim, S., S. Malek, S. Ibrahim, K. K. S., & Aziz, M. F. (2020). Risk stratification of Asian patients after ST-elevation myocardial infarction using machine learning methods. *European Heart Journal, 41*(Supplement_2), ehaa946-3494.

Kassambara, A. (2018). *Machine learning essentials: Practical Guide in R.* France: Sthda.

Katsouyanni, K., Touloumi, G., Samoli, E., Gryparis, A., Le Tertre, A., Monopolis, Y., . . . Anderson, H. (2001). Confounding and effect modification in the short-term effects of ambient particles on total mortality: results from 29 European cities within the APHEA2 project. *Epidemiology*, 521-531.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., . . . Liu, T. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, Article#30.

Ke, J., Chen, Y., Wang, X., Wu, Z., Zhang, Q., Lian, Y., & Chen, F. (2022). Machine learning-based in-hospital mortality prediction models for patients with acute coronary syndrome. *The American Journal of Emergency Medicine, 53*, 127-134.

Keim, D. A. (2005). Information visualization: Scope, techniques and opportunities for geovisualization. *Elsevier*, 21-52.

Kendall, K. E., & Kendall, J. E. (2002). *Systems analysis and design (Vol. 4).* Upper Saddle River, NJ: Prentice Hall.

Khan, A., Asif, H., Shah, A., Khan, Z., Ashraf, S., & Ashraf, A. (2022). PATIENTS PRESENTING WITH CHEST PAIN TO CARDIAC EMERGENCIES SERVICES AND APPLICATION OF TIMI RISK SCORE FOR BETTER OUTCOME. . *Pakistan Heart Journal, 55(Supplement1)*, S28-S28.

Khaniabadi, Y., Daryanoosh, S., Hopke, P., Ferrante, M., De Marco, A., Sicard, P., . . . Keishams, F. (2017). Acute myocardial infarction and COPD attributed to ambient SO2 in Iran. *Environmental research, 156*, 683-687.

Khaniabadi, Y., Polosa, R., Chuturkova, R., Daryanoosh, M., Goudarzi, G., Borgini, A., . . . Babaei, A. (2017). Human health risk assessment due to ambient PM10 and SO2 by an air quality modeling technique. *Process safety and environmental protection, 111*, 346-354.

Khennou, F., Fahim, C., Chaoui, H., & Chaoui, N. (2019). A machine learning approach: Using predictive analytics to identify and analyze high risks patients with heart disease. *International Journal of Machine Learning and Computing, 9(6)*, 762-767.

Khera, A. V., & Kathiresan, S. (2017). Genetics of coronary artery disease: discovery, biology and clinical translation. *Nature Reviews Genetics, 18(6)*, 331-344.

Khir, M. S., Muda, K., Hussein, N., Khanan, M. F., Othman, M. N., Hashim, N., & Dahari, N. (2018). Spatio-temporal analysis of PM10 in Southern Peninsular Malaysia. *International Journal of Engineering and Technology, 7(3)*, 27-30.

Kim, E., Han, K., Cheong, T., Lee, S., Eun, J., & Kim, S. (2022). Analysis on Benefits and Costs of Machine Learning-Based Early Hospitalization Prediction. *IEEE Access, 10*, 32479-32493.

Kim, J. (2017). Big data, health informatics, and the future of cardiovascular medicine. *Journal of the American College of Cardiology, 69(7)*, 899-902.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence (Vol. 2)* (pp. 1137-1143). Morgan Kaufmann Publishers Inc.

Kong, G. C., Ng, C., Muthiah, M., Foo, R., Vernon, S., Loh, P., & Chan, M. (2023). Higher mortality in acute coronary syndrome patients without standard modifiable risk factors: Results from a global meta-analysis of 1,285,722 patients. *International Journal of Cardiology, 371*, 432-440.

Koulova, A., & Frishman, W. H. (2014). Air pollution exposure as a risk factor for cardiovascular disease morbidity and mortality. *Cardiology in review, 22(1)*, 30-36.

Kranc, H., Novack, V., Shtein, A., Sonkin, R., Jaffe, E., & Novack, L. (2021). Ambient air pollution and out-of-hospital cardiac arrest. Israel nation wide assessment. *Atmospheric Environment, 261*, Article#118567.

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.

Krum, R. (2013). Cool infographics: Effective communication with data visualization and design. *John Wiley & Sons*, 2-19.

Kuhn, M., & contributors. (2023). *Caret Package*. Retrieved from The Comprehensive R Archive Network: https://cran.r-project.org/web/packages/caret/vignettes/caret.html

Kulkarni, A., Chong, D., & Batarseh, F. A. (2020). Foundations of data imbalance and solutions for a data democracy. *data democracy* (pp. 83-106). Academic Press.

Kumar, A., & Cannon, C. (2009). Acute coronary syndromes: diagnosis and management, part I. *Mayo Clinic Proceedings* (pp. 917-938). Elsevier.

Kumar, R., & Indrayan, A. (2011). Receiver operating characteristic (ROC) curve for medical researchers. *Indian pediatrics, 48*, 277-287.

Kumar, S. S., Sasidharan, A., & Bagepally, B. S. (2023). Air pollution and cardiovascular disease burden: changing patterns and implications for public health in India. *Heart, Lung and Circulation, 32(1)*, 90-94.

Kumbhani, D., Wells, B., Lincoff, A., Jain, A., Arrigain, S., Yu, C., . . . Kattan, M. (2013). Predictive models for short-and long-term adverse outcomes following discharge in a contemporary population with acute coronary syndromes . *American Journal of Cardiovascular Disease, 3(1)*, Article#39.

Kuno, T., Sahashi, Y., Kawahito, S., Takahashi, M., Iwagami, M., & Egorova, N. N. (2022). Prediction of in-hospital mortality with machine learning for COVID-19 patients treated with steroid and remdesivir. *Journal of Medical Virology, 94(3)*, 958-964.

Kuźma, Ł., Pogorzelski, S., Struniawski, K., Dobrzycki, S., & Bachórzewska-Gajewska, H. (2019). Effect of air pollution on the number of hospital admissions for acute coronary syndrome in elderly patients. *Polish Archives of Internal Medicine, 130(1)*, 38-46.

Kuźma, Ł., Wańha, W., Kralisz, P., Kazmierski, M., Bachórzewska-Gajewska, H., Wojakowski, W., & Dobrzycki, S. (2021). Impact of short-term air pollution exposure on acute coronary syndrome in two cohorts of industrial and non-industrial areas: A time series regression with 6,000,000 person-years of follow-up (ACS-Air Pollution Study). *Environmental Research, 197*, Article#111154.

Kwon, H., Park, J., & Lee, Y. (2019). Stacking ensemble technique for classifying breast cancer. *Healthcare informatics research, 25(4)*, 283-288.

Kwon, J.-m., Jeon, K.-H., Kim, H. M., Kim, M. J., Lim, S., Kim, K.-H., . . . Oh, B.-H. (2019a). Deep-learning-based risk stratification for mortality of patients with acute myocardial infarction. *Plos One, 14(10)*, Article#e0224502.

Laden, F., Schwartz, J., Speizer, F., & Dockery, D. (2006). Reduction in fine particulate air pollution and mortality: extended follow-up of the Harvard Six Cities study. *American journal of respiratory and critical care medicine, 173(6)*, 667-672.

Lall, R., Kendall, M., Ito, K., & Thurston, G. D. (2004). Estimation of historical annual PM2. 5 exposures for health effects assessment. *Atmospheric Environment, 38(31)*, 5217-5226.

Larsen, M. P., Eisenberg, M. S., Cummins, R. O., & Hallstrom, A. P. (1993). Predicting survival from out-of-hospital cardiac arrest: a graphic model. *Annals of emergency medicine, 22(11)*, 1652-1658.

Lau, C. F., Malek, S., Gunalan, R., Chee, W. H., Saw, A., & Aziz, F. (2022). Paediatric upper limb fracture healing time prediction using a machine learning approach. *All Life, 15(1)*, 490-499.

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)*, Article#e0127428.

Lee, W., Lee, J., Woo, S., Choi, S., Bae, J., Jung, S., . . . Lee, W. (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)*, Article#12886.

Leem, J. H., Kim, S. T., & Kim, H. C. (2015). Public-health impact of outdoor air pollution for 2nd air pollution management policy in Seoul metropolitan area, Korea. *Annals of occupational and environmental medicine, 27*, 1-9.

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., & Pozzer, A. (2015). The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature, 525(7569)*, 367-371.

Leung, K. M. (2007). Naive bayesian classifier. *Polytechnic University Department of Computer Science/Finance and Risk Engineering, 2007*, 123-156.

Lewis, J. R. (2018). The system usability scale: past, present, and future. *International Journal of Human–Computer Interaction, 34(7)*, 577-590.

Li, H., Cai, J., Chen, R., Zhao, Z., Ying, Z., Wang, L., . . . Kan, H. (2017). Particulate matter exposure and stress hormone levels: a randomized, double-blind, crossover trial of air purification. *Circulation, 136(7)*, 618-627.

Li, M., Fan, L., Mao, B., Yang, J., Choi, A., Cao, W., & Xu, J. (2016). Short-term exposure to ambient fine particulate matter increases hospitalizations and mortality in COPD: a systematic review and meta-analysis. *Chest, 149(2)*, 447-458.

Li, M., Fu, X., & Li, D. (2020). Diabetes prediction based on XGBoost algorithm. In IOP conference series: materials science and engineering (Vol. 768, No. 7) . *IOP Publishing.*, Article#072093.

Liao, Z., Gao, M., Sun, J., & Fan, S. (2021). The impact of synoptic circulation and long-term circulation change on air quality and pollution-related human health in the Yangtze River Delta region. *Air Pollution, Climate, and Health* (pp. 135-161). Elsevier.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news, 2(3)*, 18-22.

Lin, K., & Gao, Y. (2022). Model interpretability of financial fraud detection by group SHAP. *Expert Systems with Applications, 210*, Article#118354.

Lin, Y. C., Tsai, C. H., Hsu, H. T., & Lin, C. H. (2021). Using machine learning to analyze and predict the relations between cardiovascular disease incidence, extreme temperature and air pollution. *2021 IEEE 3rd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS)* (pp. 234-237). IEEE.

Ling, C. X., Huang, J., & Zhang, H. (2003). AUC: a better measure than accuracy in comparing learning algorithms. *Advances in Artificial Intelligence: 16th Conference of the Canadian Society for Computational Studies of Intelligence, AI 2003* (pp. 329-341). Canada: Springer.

Lippi, G., Franchini, M., Montagnana, M., Filippozzi, L., Favaloro, E., & Guidi, G. (2010). Relationship between 24-h air pollution, emergency department admission and diagnosis of acute coronary syndrome. *Journal of thrombosis and thrombolysis, 29*, 381-386.

Liu, C., Chen, R., Meng, X., Wang, W., Lei, J., Zhu, Y., . . . Xuan, J. (2022). Criteria air pollutants and hospitalizations of a wide spectrum of cardiovascular diseases: A nationwide case-crossover study in China. *Eco-Environment & Health, 1(4)*, 204-211.

Liu, Y., Liu, Z., Luo, X., & Zhao, H. (2022). Diagnosis of Parkinson's disease based on SHAP value feature selection. *Biocybernetics and Biomedical Engineering, 42(3)*, 856-869.

Louridi, N., Amar, M., & El Ouahidi, B. (2019). Identification of cardiovascular diseases using machine learning. *IEEE*, 1-6.

Lu, H., & Nordin, R. (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.

Lu, X., Lin, C., Li, W., Chen, Y., Huang, Y., Fung, J., & Lau, A. (2019). Analysis of the adverse health effects of PM2. 5 from 2001 to 2017 in China and the role of urbanization in aggravating the health burden. *Science of the Total Environment, 652*, 683-695.

Lunardon, N., Menardi, G., & Torelli, N. (2014). ROSE: a package for binary imbalanced learning. *R journal, 6(1)*, 79-89.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems, 30*, 1-10.

Lundberg, S., Nair, B., Vavilala, M., Horibe, M., Eisses, M., Adams, T., . . . Lee, S. (2018). Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature biomedical engineering, 2(10)*, 749-760.

Maheswari, S., & Pitchai, R. (2019). Heart disease prediction system using decision tree and naive Bayes algorithm. *Current Medical Imaging, 15(8)*, 712-717.

Makmom Abdullah, A., Armi Abu Samah, M., & Yee Jun, T. (2012). An overview of the air pollution trend in Klang Valley, Malaysia. *Open Environmental Sciences, 6(1)*, 13-19.

Manisalidis, I., Stavropoulou, E., Stavropoulos, A., & Bezirtzoglou, E. (2020). Environmental and health impacts of air pollution: a review. *Frontiers in public health*, Article#14.

Mansoor, H., & Bian, J. (2017). Risk Prediction for in-Hospital Mortality in Women With St-Elevation Myocardial Infarction: a Machine Learning Approach. *Journal of the American College of Cardiology, 69(11)*, Article#171.

Mansoor, H., Elgendy, I., Segal, R., Bavry, 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.

Martin, M. (2023, April 9). *Prototype Model in Software Engineering*. Retrieved from Guru99: https://www.guru99.com/software-engineering-prototyping-model.html

Martinez-Sanchez, C., Borrayo, G., Carrillo, J., Juarez, U., Quintanilla, J., & Jerjes-Sanchez, C. (2016). Clinical management and hospital outcomes of acute coronary syndrome patients in Mexico: The Third National Registry of Acute Coronary Syndromes (RENASICA III). *Archivos de cardiología de México, 86(3)*, 221-232.

Martins, L., Pereira, L., Lin, C., Santos, U., Prioli, G., Luiz, O., . . . Braga, A. (2006). The effects of air pollution on cardiovascular diseases: lag structures. *Revista de Saúde Pública, 40(4)*, 677-683.

Masetic, Z., & Subasi, A. (2016). Congestive heart failure detection using random forest classifier. *Computer methods and programs in biomedicine, 130*, 54-64.

Masmuzidin, M., & Aziz, N. (2019). The adaptation of Shneiderman's golden rules and nielsen's heuristics on motivational augmented reality technology design for young children. *2019 IEEE 9th International Conference on System Engineering and Technology (ICSET)* (pp. 62-67). IEEE.

Matheny, M., McPheeters, M. L., Glasser, A., Mercaldo, N., Weaver, R. B., Jerome, R. N., . . . Tsai, C. (2011). Systematic review of cardiovascular disease risk assessment tools. *gency for Healthcare Research and Quality (US), Rockville (MD)*, Article#21796824 .

Mazeli, M. I., Pahrol, M. A., Shakor, A. S., Kanniah, K. D., & Omar, M. A. (2023). Cardiovascular, respiratory and all-cause (natural) health endpoint estimation using a spatial approach in Malaysia. *Science of The Total Environment, 874*, Article#162130.

Mazeli, M. I., Pahrol, M. A., Shakor, A. S., Kanniah, K. D., & Omar, M. A. (2023). Cardiovascular, respiratory and all-cause (natural) health endpoint estimation using a spatial approach in Malaysia. *Science of The Total Environment, 874*, Article#162130.

McGranaghan, M. (1993). A cartographic view of spatial data quality. *Cartographica: The International Journal for Geographic Information and Geovisualization, 30(2-3)*, 8-19.

Meng, X., Zhang, Y., Yang, K. Q., Yang, Y. K., & Zhou, X. L. (2016). Potential harmful effects of PM2. 5 on occurrence and progression of acute coronary syndrome: epidemiology, mechanisms, and prevention measures. *International journal of environmental research and public health, 13(8)*, Article#748.

Mezzatesta, S., Torino, C., De Meo, P., Fiumara, G., & Vilasi, A. (2019). A machine learning-based approach for predicting the outbreak of cardiovascular diseases in patients on dialysis. *Computer methods and programs in biomedicine, 177*, 9-15.

Microsoft. (2016). *Welcome to LightGBM's documentation*. Retrieved from LightGBM 3.3.2 documentation: https://lightgbm.readthedocs.io/en/v3.3.2/

Miller, K. A., Siscovick, D. S., Sheppard, L., Shepherd, K., Sullivan, J. H., Anderson, G. L., & Kaufman, J. D. (2007). Long-term exposure to air pollution and incidence of cardiovascular events in women. *New England Journal of Medicine, 356(5)*, 447-458.

Ministry of Health Malaysia, M. (2017). *Cardiovascular diseases clinical practice guidelines*. Retrieved from Ministry of Health Malaysia: http://www.moh.gov.my/moh/resources/Penerbitan/CPG/CARDIOVASCULAR/3.pdf

Miranda, I., Cardoso, G., Pahar, M., Oliveira, G., & Niesler, T. (2021). Machine learning prediction of hospitalization due to COVID-19 based on self-reported symptoms: A study for Brazil. *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)* (pp. 1-5). IEEE.

Mirza, A., Taha, A., & Khdhir, B. (2018). Risk factors for acute coronary syndrome in patients below the age of 40 years. *The Egyptian Heart Journal, 70(4)*, 233-235.

Morrow, D., Antman, E., Charlesworth, A., Cairns, R., Murphy, S., de Lemos, J., . . . 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.

Murad, M. H. (2012). Main air pollutants and myocardial infarction. *J. Am. Med. Assoc. Rev, 307*, 713-721.

Nadakinamani, R. G., Reyana, A., Kautish, S., Vibith, A. S., Gupta, Y., Abdelwahab, S. F., & Mohamed, A. W. (2022). Clinical data analysis for prediction of cardiovascular disease using machine learning techniques. *Computational Intelligence and Neuroscience*, 1-13.

Nag, T., & Ghosh, A. (2013). Cardiovascular disease risk factors in Asian Indian population: A systematic review. *Journal of cardiovascular disease research, 4(4)*, 222-228.

Naimish, S. (2023). *Software prototyping model and phases*. Retrieved from GeeksforGeeks: https://www.geeksforgeeks.org/software-prototyping-model-and-phases/

Nakanishi, R., Slomka, P., Rios, R., Betancur, J., Blaha, M., Nasir, K., . . . Rozanski, A. (2021). Machine learning adds to clinical and CAC assessments in predicting 10-year CHD and CVD deaths. *Cardiovascular Imaging, 14(3)*, 615-625.

Nansseu, J., Alima Yanda, A., Chelo, D., Tatah, S., Mbassi Awa, H., Seungue, J., & Koki, P. (2015). The Acute Chest Syndrome in Cameroonian children living with sickle cell disease. *BMC pediatrics, 15(1)*, 1-8.

Natarajan, P. (2018). Polygenic risk scoring for coronary heart disease: the first risk factor. *Journal of the American College of Cardiology, 72(16)*, 1894-1897.

Neves, V., Roman, R., Vendruscolo, T., Heineck, G., Mattos, C., Mattos, E., . . . Roman, M. (2021). Validation of the Grace Risk Score to Predict In-Hospital and 6-Month Post-Discharge Mortality in Patients with Acute Coronary Syndrome. *International Journal of Cardiovascular Sciences, 35*, 174-180.

Nguyen, T., Vo, P., Huynh, H., Tran, A., Tran, C., Vi, M., . . . Taxis, K. (2021). Performance of the GRACE 2.0 and EPICOR risk scores for predicting 1-year postdischarge mortality in Vietnamese patients with acute coronary syndrome. *Pharmaceutical Sciences Asia, 48(4)*, 367-374.

Nirel, R., & Dayan, U. (2001). On the ratio of sulfur dioxide to nitrogen oxides as an indicator of air pollution sources. *Journal of Applied Meteorology, 40(7)*, 1209-1222.

Nogueira, J. B. (2009). Air pollution and cardiovascular disease. *Revista portuguesa de cardiologia: orgao oficial da Sociedade Portuguesa de Cardiologia= Portuguese* *journal of cardiology: an official journal of the Portuguese Society of Cardiology,* , 28(6).

Norwegian Research Council, N. (2017). *baseline characteristics*. Retrieved from GET-IT Glossary: https://getitglossary.org/term/baseline+characteristics

Nuckols, J. R., Ward, M. H., & Jarup, L. (2004). Using geographic information systems for exposure assessment in environmental epidemiology studies. *Environmental health perspectives, 112(9)*, 1007-1015.

Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—big data, machine learning, and clinical medicine. *The New England Journal of Medicine, 375(13)*, Article#1216.

O'gara, P., Kushner, F., Ascheim, D., Casey, D., Chung, M., De Lemos, J., . . . Granger, C. (2013). 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Journal of the American College of Cardiology, 61(4)*, e78-e140.

Ohira, T., & Iso, H. (2013). Cardiovascular disease epidemiology in Asia–an overview–. *Circulation Journal, 77(7)*, 1646-1652.

Ono, M., Kawashima, H., Hara, H., Gamal, A., Wang, R., Gao, C., . . . Jüni, P. (2021). External validation of the GRACE risk score 2.0 in the contemporary all-comers GLOBAL LEADERS trial. *Catheterization and cardiovascular interventions, 98(4)*, E513-E522.

O'Toole, T. E., Conklin, D. J., & Bhatnagar, A. (2008). Environmental risk factors for heart disease. *Reviews on environmental health, 23(3)*, 167-202.

Overbaugh, K. J. (2009). Acute coronary syndrome. *AJN The American Journal of Nursing, 109(5)*, 42-52.

Özen, A., Gönen, M., Alpaydın, E., & Haliloğlu, T. (2009). Machine learning integration for predicting the effect of single amino acid substitutions on protein stability. *BMC Structural Biology, 9(1)*, Article#66.

Palacio-Niño, J. O., & Berzal, F. (2019). Evaluation metrics for unsupervised learning algorithms. *arXiv preprint arXiv*, 1905.05667.

Parab, J., Sequeira, M., Lanjewar, M., Pinto, C., & Naik, G. (2021). Backpropagation neural network-based machine learning model for prediction of blood urea and glucose in CKD patients. *IEEE journal of translational engineering in health and medicine, 9*, 1-8.

Park, H., & Kim, S. (2021). Hardware accelerator systems for artificial intelligence and machine learning. *Advances in Computers (Vol. 122)* (pp. 51-95). Elsevier.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research, 12*, 2825-2830.

Pelucchi, C., Negri, E., Gallus, S., Boffetta, P., Tramacere, I., & La Vecchia, C. (2009). Long-term particulate matter exposure and mortality: a review of European epidemiological studies. *BMC public health, 9(1)*, 1-8.

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.

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.

Peng, Y., Xin, D., Kris D., R., Yang Feng, W., Runlin, G., & Anushka, P. (2017). Predicting In-Hospital Mortality in Patients with Acute Coronary Syndrome in China. *The American journal of cardiology 120, no. 7*, 1077-1083.

Peter J., B., Aaron, Y., Aarjav, J., Nicholas, P., Sergio, V., & Solomon, F. B. (2022). Gaussian-Process based inference of electrolyte decomposition reaction networks in Li-ion battery failure. In M. Ludovic, & N. Stephane, *Computer Aided Chemical Engineering* (pp. 157-162). Elsevier.

Peterson, P. L., Baker, E., & McGaw, B. (2010). *International encyclopedia of education.* Elsevier Ltd.

Pieper, K., Gore, J., FitzGerald, G., Granger, C., Goldberg, R., Steg, G., . . . Global Registry of Acute Coronary Events (GRACE) Investigators. (2009). Validity of a risk-prediction tool for hospital mortality: the Global Registry of Acute Coronary Events . *American heart journal, 157(6)*, 1097-1105.

Pinto, M., Marotta, N., Caracò, C., Simeone, E., Ammendolia, A., & de Sire, A. (2022). Quality of life predictors in patients with melanoma: a machine learning approach. *Frontiers in Oncology, 12*, 843611.

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.

Pope CA, 3., & Dockery, D. (2006). Health effects of fine particulate air pollution: lines that connect. *J Air Waste Manag Assoc. 6(6)*, 709–742.

Pope III, C., Muhlestein, J., Anderson, J., Cannon, J., Hales, N., Meredith, K., . . . Horne, B. (2015). Short-term exposure to fine particulate matter air pollution is preferentially associated with the risk of ST-segment elevation acute coronary events. *Journal of the American Heart Association, 4(12)*, p.e002506.

Pope, C., Brook, R., Burnett, R., & Dockery, D. (2011). How is cardiovascular disease mortality risk affected by duration and intensity of fine particulate matter exposure? An integration of the epidemiologic evidence. *Air Quality, Atmosphere & Health, 4*, 5-14.

Prattichizzo, F., de Candia, P., De Nigris, V., Nicolucci, A., & Ceriello, A. (2020). Legacy effect of intensive glucose control on major adverse cardiovascular outcome:

systematic review and meta-analyses of trials according to different scenarios. *Metabolism, 110*, Article#154308.

Psychogyios, K., Ilias, L., & Askounis, D. (2022). Comparison of Missing Data Imputation Methods using the Framingham Heart study dataset. *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)* (pp. 1-5). IEEE.

Qi, Y. (2012). Random Forest for Bioinformatics. *Ensemble Machine Learning: Methods and Applications* (pp. 307-323). Boston: Springer.

Qiu, H., Luo, L., Su, Z., Zhou, L., Wang, L., & Chen, Y. (2020). Machine learning approaches to predict peak demand days of cardiovascular admissions considering environmental exposure. *BMC medical informatics and decision making, 20(1)*, 1-11.

Qiu, X., Wei, Y., Wang, Y., Di, Q., Sofer, T., Awad, Y. A., & Schwartz, J. (2020). Inverse probability weighted distributed lag effects of short-term exposure to PM2.5 and ozone on CVD hospitalizations in New England Medicare participants—Exploring the casual effects. *Environmental Research, 182*, Article#109095.

Qorbani, M., Yunesian, M., Fotouhi, A., Zeraati, H., & Sadeghian, S. (2012). Effect of air pollution on onset of acute coronary syndrome in susceptible subgroups. *EMHJ-Eastern Mediterranean Health Journal, 18 (6)*, 550-555.

Radack, S. (2009). *System Development Life Cycle*. Retrieved from National Institute of Standards and Technology: https://csrc.nist.gov/csrc/media/publications/shared/documents/itlbulletin/itlbul2009-04.pdf

Radović, N., Prelević, V., Erceg, M., & Antunović, T. (2022). Machine learning approach in mortality rate prediction for hemodialysis patients. *Computer Methods in Biomechanics and Biomedical Engineering, 25(1)*, 111-122.

Rajak, R., & Chattopadhyay, A. (2020). Short and long term exposure to ambient air pollution and impact on health in India: a systematic review. *International journal of environmental health research, 30(6)*, 593-617.

Ralapanawa, U., Kumarasiri, P., Jayawickreme, K., 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.

Rana, D. (2015). One class SVM vs SVM classification. *Int. J. Sci. Res., 4(6)*, 1350-1352.

Rani, N. L., Azid, A., Khalit, S. I., Juahir, H., & Samsudin, M. S. (2018). Air Pollution Index Trend Analysis in Malaysia, 2010-15. *Polish Journal of Environmental Studies, 27(2)*, 801 - 807.

Ranjith, N., Pegoraro, R., & Naidoo, D. (2005). Demographic data and outcome of acute coronary syndrome in the South African Asian Indian population: Cardiovascular topic. *Cardiovascular Journal of South Africa, 16(1)*, 48-54.

Rao, S., & Agasthi, P. (2023, February 6). *Thrombolysis In Myocardial Infarction Risk Score*. Retrieved from Treasure Island (FL): StatPearls Publishing: https://www.ncbi.nlm.nih.gov/books/NBK556069/

Rao, S., Mehta, S., Kulkarni, S., Dalvi, H., Katre, N., & Narvekar, M. (2022). A Study of LIME and SHAP Model Explainers for Autonomous Disease Predictions. *2022 IEEE Bombay Section Signature Conference (IBSSC)* (pp. 1-6). Bombay : IEEE.

Ravindra, K., Bahadur, S. S., Katoch, V., Bhardwaj, S., Kaur-Sidhu, M., Gupta, M., & Mor, S. (2023). Application of machine learning approaches to predict the impact of ambient air pollution on outpatient visits for acute respiratory infections. *Science of The Total Environment, 858*, Article#159509.

Raza, A., Bellander, T., Bero-Bedada, G., Dahlquist, M., Hollenberg, J., Jonsson, M., . . . Ljungman, P. (2014). Short-term effects of air pollution on out-of-hospital cardiac arrest in Stockholm. *European heart journal, 35(13)*, 861-868.

Redpath, D. B., & Lebart, K. (2005). Boosting feature selection. *International Conference on Pattern Recognition and Image Analysis* (pp. 305-314). Berlin: Springer.

Reigle, J. (2005). Evaluating the patient with chest pain: the value of a comprehensive history. *Journal of Cardiovascular Nursing, 20(4)*, 226-231.

Ren, H., Sun, Y., Xu, C., Fang, M., Xu, Z., Jing, F., . . . Jin, W. (2022). Predicting Acute Onset of Heart Failure Complicating Acute Coronary Syndrome: an Explainable Machine Learning Approach. *Current problems in cardiology*, Article#101480.

Renukadevi, N. T., & P., T. (2013). Performance evaluation of SVM–RBF kernel for medical image classification. *Global Journal of Computer Science and Technology* , 1.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should i trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144). Association for Computing Machinery.

Rigatti, S. J. (2017). Random forest. *Journal of Insurance Medicine, 47(1)*, 31-39.

Robert, K. W., Parris, T. M., & Leiserowitz, A. A. (2005). What is sustainable development? Goals, indicators, values, and practice. *Environment: science and policy for sustainable development, 47(3)*, 8-21.

Rodríguez-Pérez, R., & Bajorath, J. (2019). Interpretation of compound activity predictions from complex machine learning models using local approximations and shapley values. *Journal of medicinal chemistry, 63(16)*, 8761-8777.

Roshan, A., Choo, J., & Lim, C. (2023). Readability, Understandability, and Actionability of Online Cardiovascular Risk Assessment Tools and Patient Educational Material: A Systematic Review. *Glomerular Diseases, 3*, 56-68.

Roth, G. (2018). Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2017 (GBD 2017) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2018. *The Lancet, 392*, 1736-1788.

Roth, G., Mensah, G., Johnson, C., Addolorato, G., Ammirati, E., Baddour, L., . . . Bonny, A. (2020). Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study. *Journal of the American College of Cardiology, 76(25)*, 2982-3021.

Rukshan, P. (2020, December 19). *k-fold cross-validation explained in plain English*. Retrieved from Towards Data Science: https://towardsdatascience.com/k-fold-cross-validation-explained-in-plain-english-659e33c0bc0

Rus, A.-A., & Mornoş, C. (2022). The Impact of Meteorological Factors and Air Pollutants on Acute Coronary Syndrome. . *Current Cardiology Reports, 24(10)*, 1337–1349.

Ruvira, G., Ruvira-Durante, J., Cosín-Sales, J., Marín-García, P. J., & Llobat, L. (2023). Environmental gaseous pollutants are related to increase of acute coronary syndrome in Valencia region. *Medicina Clínica*, 1-5.

Saar, A., Marandi, T., Ainla, T., Fischer, K., Blöndal, M., & Eha, J. (2018). The risk-treatment paradox in non-ST-elevation myocardial infarction patients according to their estimated GRACE risk. *International journal of cardiology, 272*, 26-32.

Sallehuddin, H. M., Azman, S. F., Noor, S. M., Faisal, U. A., Pillai, S. V., & Aziz, A. A. (2017). Global Registry of Acute Coronary Events (GRACE) Risk Score in Predicting Outcome in Elderly Patients with ST Elevation Myocardial Infarction at 6 Months After Primary Percutaneous Coronary Intervention in Hospital Serdang. *International Journal of Cardiology, 249,* , S32-S33.

Samet, J. M., Dominici, F., Curriero, F. C., Coursac, I., & Zeger, S. L. (2000). Fine particulate air pollution and mortality in 20 US cities, 1987–1994. *New England journal of medicine, 343(24)*, 1742-1749.

Sammut, C., & Webb, G. I. (2011). *Encyclopedia of machine learning.* Springer Science & Business Media.

Samoli, E., Peng, R., Ramsay, T., Pipikou, M., Touloumi, G., Dominici, F., . . . Katsouyanni, K. (2008). Acute effects of ambient particulate matter on mortality in Europe and North America: results from the APHENA study. . *Environmental health perspectives, 116(11)*, 1480-1486.

Santurtún, A., Sanchez-Lorenzo, A., Villar, A., Riancho, J. A., & Zarrabeitia, M. T. (2017). The Influence of Nitrogen Dioxide on Arrhythmias in Spain and Its Relationship with Atmospheric Circulation. *Cardiovascular Toxicology, 17(1)*, 88–96.

Saraswat, M. (2016, December). *Beginners Tutorial on XGBoost and Parameter Tuning in R*. Retrieved from hackerearth: https://www.hackerearth.com/practice/machine-learning/machine-learning-algorithms/beginners-tutorial-on-xgboost-parameter-tuning-r/tutorial/

Saritas, M. M., & Yasar, A. (2019). Performance analysis of ANN and Naive Bayes classification algorithm for data classification. *International journal of intelligent systems and applications in engineering, 7(2)*, 88-91.

Seliya, N., Khoshgoftaar, T. M., & Van Hulse, J. (2009). A study on the relationships of classifier performance metrics. *2009 21st IEEE international conference on tools with artificial intelligence* (pp. 59-66). IEEE.

Selvarajah, S., Fong, A. 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)*, Article#e40249.

Selvarajah, S., Fong, A., Selvaraj, G., Haniff, J., Hairi, N., Bulgiba, A., & Bots, M. (2013). Impact of cardiac care variation on ST-elevation myocardial infarction outcomes in Malaysia. *The American journal of cardiology, 111(9)*, 1270-1276.

Serafeim, L. (2020, May 28). *Everything you need to know about Min-Max normalization: A Python tutorial*. Retrieved from Towards Data Science: https://towardsdatascience.com/everything-you-need-to-know-about-min-max-normalization-in-python-b79592732b79

Shailaja, K., Seetharamulu, B., & Jabbar, M. A. (2018). Machine learning in healthcare: A review. *2018 Second international conference on electronics, communication and aerospace technology (ICECA)* (pp. 910-914). IEEE.

Shaji, S., Palanisamy, R., & Swaminathan, R. (2022). Explainable Optimized LightGBM Based Differentiation of Mild Cognitive Impairment Using MR Radiomic Features. *Studies in Health Technology and Informatics, 295*, 483-486.

Shang, Y., Sun, Z., Cao, J., Wang, X., Zhong, L., Bi, X., . . . Huang, W. (2013). Systematic review of Chinese studies of short-term exposure to air pollution and daily mortality. *Environment international, 54*, 100-111.

SHAP, G. (2017). *Welcome to the SHAP documentation - SHAP latest documentation.* . Retrieved from Github: https://shap.readthedocs.io/en/latest/index.html

Sharma, D. K., Chatterjee, M., Kaur, G., & Vavilala, S. (2022). Deep learning applications for disease diagnosis. *Deep learning for medical applications with unique data* (pp. 31-51). Academic Press.

Sherazi, S. W., 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)*, Article#e0249338.

Shneiderman, B. (1986). Eight golden rules of interface design. *Disponible en*, 172.

Shneiderman, B., & Plaisant, C. (2004). *Designing the user interface: Strategies for effective Human-Computer Interaction (4th ed.).* Boston: MA: Addison Wesley.

Shouval, R., Hadanny, A., Shlomo, N., Iakobishvili, Z., Unger, R., Zahger, D., . . . Goldenberg, I. (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.

Sia, C., Zheng, H., Ko, J., Ho, A., Foo, D., Foo, L., . . . Tan, H. (2022). Comparison of the modified Singapore myocardial infarction registry risk score with GRACE 2.0 in predicting 1-year acute myocardial infarction outcomes. *Scientific Reports, 12(1)*, Article#14270.

Sidhu, N. S., Rangaiah, S. K., Ramesh, D., Veerappa, K., & Manjunath, C. N. (2020). Clinical characteristics, management strategies, and in-hospital outcomes of acute coronary syndrome in a low socioeconomic status cohort: an observational study from urban India. *Clinical Medicine Insights: Cardiology, 14*, 1179546820918897.

Sihwi, S. W., Jati, I. P., & Anggrainingsih, R. (2018). Twitter sentiment analysis of movie reviews using information gain and naïve bayes classifier. *2018 International Seminar on Application for Technology of Information and Communication* (pp. 190-195). IEEE.

Silberbauer, M. a. (2009). Google Earth: A spatial interface for SA water resource data. *PositionIT April/May 2009*, 42-47.

Simkhovich, B. Z., Kleinman, M. T., & Kloner, R. A. (2008). Air pollution and cardiovascular injury: epidemiology, toxicology, and mechanisms. *Journal of the american college of cardiology, 52(9)*, 719-726.

Simske, S. (2019). *Meta-analytics: consensus approaches and system patterns for data analysis.* Morgan Kaufmann.

Singh, P., Singh, N., Singh, K. K., & Singh, A. (2021). Diagnosing of disease using machine learning. *Machine learning and the internet of medical things in healthcare* (pp. 89-111). Academic Press.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing, 14(3)*, 199-222.

Soiza, R., Leslie, S., Harrild, K., Peden, N., & Hargreaves, A. (2005). Age-dependent differences in presentation, risk factor profile, and outcome of suspected acute coronary syndrome. *Journal of the American Geriatrics Society, 53(11)*, 1961-1965.

Song, L., Langfelder, P., & Horvath, S. (2013). Random generalized linear model: a highly accurate and interpretable ensemble predictor. *BMC bioinformatics, 14(1)*, 1-22.

Song, Y., Jiao, X., Yang, S., Zhang, S., Qiao, Y., Liu, Z., & Zhang, L. (2019). Combining multiple factors of LightGBM and XGBoost algorithms to predict the morbidity of double-high disease. *International Conference of Pioneering Computer Scientists, Engineers and Educators* (pp. 635-644). Singapore: Springer.

Sposito, A., & Chapman, M. (2002). Arteriosclerosis, thrombosis, and vascular biology, 22(10). *Statin therapy in acute coronary syndromes: mechanistic insight into clinical benefit*, 1524-1534.

Stanaway, J., Afshin, A., Gakidou, E., Lim, S., Abate, D., Abate, K., . . . Abdela, J. (2018). Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017. *Lancet (London, England) 392, no. 10159*, 1923-1994.

Starovoitov, V., & Golub, Y. (2021). Data normalization in machine learning. *Informatics*, 244221434.

Stefanou, C. (2003). System Development Life Cycle. In H. Bidgoli, *Encyclopedia of Information Systems* (pp. 329-344). Elsevier.

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., . . . Matama, H. (2021). Cardiac outcomes in patients with acute coronary syndrome attributable to calcified nodule. *Atherosclerosis, 318*, 70-75.

Sun, H., Burton, H. V., & Huang, H. (2021). Machine learning applications for building structural design and performance assessment: State-of-the-art review. *Journal of Building Engineering, 33*, Article#101816.

Sun, Y., & Zhou, Y. H. (2022). A Machine Learning Pipeline for Mortality Prediction in the ICU. *International Journal of Digital Health, 2(1)*, Article#3.

Suzuki, S., Yamashita, T., Sakama, T., Arita, T., Yagi, N., Otsuka, T., . . . Uejima, T. (2019). Comparison of risk models for mortality and cardiovascular events between machine learning and conventional logistic regression analysis. *PLoS One, 14(9)*, Article#e0221911.

Swapna, M., Viswanadhula, U. M., Aluvalu, R., Vardharajan, V., & Kotecha, K. (2022). Bio-signals in medical applications and challenges using artificial intelligence. *Journal of Sensor and Actuator Networks, 11(1)*, Article#17.

Swee-Hock, S. (2015). *The population of Malaysia (Vol. 514).* Institute of Southeast Asian Studies.

Szabó, G. T., Ágoston, A., Csató, G., Rácz, I., Bárány, T., Uzonyi, G., . . . Édes, I. F. (2021). Predictors of Hospital Mortality in Patients with Acute Coronary Syndrome Complicated by Cardiogenic Shock. *Sensors, 21(3)*, Article#969.

Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool . *BMC Medical Imaging, 15(1)*, 1-28.

Talabis, M., McPherson, R., Miyamoto, I., Martin, J., & Kaye, D. (2015). Analytics defined. *Information Security Analytics*, 1-12.

Talmor-Barkan, Y., Bar, N., Shaul, A., Shahaf, N., Godneva, A., Bussi, Y., . . . Arow, Z. (2022). Metabolomic and microbiome profiling reveals personalized risk factors for coronary artery disease. *Nature medicine, 28(2)*, 295-302.

Tang, E., Wong, C., & Herbison, P. (2007). Global Registry of Acute Coronary Events (GRACE) hospital discharge risk score accurately predicts long-term mortality post acute coronary syndrome. *American heart journal, 153(1)*, 29-35.

Tina, R. P., & Sherekar, S. S. (2013). Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification. *International Journal Of Computer Science And Applications* , 256-261.

Tran, K. (2021, September 24). *SHAP: Explain Any Machine Learning Model in Python*. Retrieved from Towards Data Science: https://towardsdatascience.com/shap-explain-any-machine-learning-model-in-python-24207127cad7

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)*, Article#e25000.

TutorialsPoints. (2023). *SDLC - Overview, Tutorials Point*. Retrieved from Tutorials Point: https://www.tutorialspoint.com/sdlc/sdlc_overview.htm

Ueshima, H., Sekikawa, A., Miura, K., Turin, T., Takashima, N., Kita, Y., . . . Nakamura, Y. (2008). Cardiovascular disease and risk factors in Asia: a selected review. *Circulation, 118(25)*, 2702-2709.

Urrea, C., & Venegas, D. (2020). Automatized follow-up and alert system for patients with chronic hypertension. *Health informatics journal, 26(4)*, 2625-2636.

Usmani, R. S., Pillai, T. R., Hashem, I. A., Marjani, M., Shaharudin, R., & Latif, M. T. (2021). Air pollution and cardiorespiratory hospitalization, predictive modeling, and analysis using artificial intelligence techniques. *Environmental Science and Pollution Research, 28(40)*, 56759-56771.

Usmani, R. S., Saeed, A., Abdullahi, A. M., Pillai, T. R., Jhanjhi, N. Z., & Hashem, I. A. (2020). Air pollution and its health impacts in Malaysia: a review. *Air Quality, Atmosphere & Health, 13*, 1093-1118.

Van Den Berg, P., & Body, R. (2018). The HEART score for early rule out of acute coronary syndromes in the emergency department: a systematic review and meta-analysis. *European Heart Journal: Acute Cardiovascular Care, 7(2)*, 111-119.

van der Sangen, N., Azzahhafi, J., Yin, D., Peper, J., Rayhi, S., Walhout, R., . . . van Bommel, R. (2022). External validation of the GRACE risk score and the risk–treatment paradox in patients with acute coronary syndrome. *Open Heart, 9(1)*, Article#e001984.

VanHouten, J., Starmer, J., Lorenzi, N., Maron, D., & Lasko, T. (2014). Machine learning for risk prediction of acute coronary syndrome. *AMIA Annual Symposium Proceedings (Vol. 2014)* (p. 1940). American Medical Informatics Association.

Vapnik, V., Guyon, I., & Hastie, T. (1995). Support vector machines. *Mach. Learn, 20*(3), 273-297.

Varghese, T., & Kumar, A. (2019). Predisposing risk factors of acute coronary syndrome (ACS): A mini review. *Journal of Pharmaceutical Sciences and Research, 11(5)*, 1999-2002.

Vembandasamy, K., Sasipriya, R., & Deepa, E. (2015). Heart diseases detection using Naive Bayes algorithm. *International Journal of Innovative Science, Engineering & Technology, 2(9)*, 441-444.

Vernon, S. T., Coffey, S., D'Souza, M., Chow, C. K., Kilian, J., Hyun, K., . . . Brieger, D. (2019). ST-Segment–Elevation Myocardial Infarction (STEMI) patients without standard modifiable cardiovascular risk factofactors—How common are they, and what are their outcomes? *Journal of the American Heart Association, 8(21)*, Article#e013296.

Veropoulos, K., Campbell, C., & Cristianini, N. (1999). Controlling the sensitivity of support vector machines. *Proceedings of the international joint conference on AI (Vol. 55)* (p. 60). Stockholm: Bristol University, United Kingdom.

Vij, R. (2023, June 6). *Machine Learning Algorithms Part 1: Linear Regression. Predict the price of diamonds using linear regression*. Retrieved from Towards Data Science: https://towardsdatascience.com/machine-learning-algorithms-part-1-linear-regression-a7079238edc9

Walford, N. S. (2020). Demographic and social context of deaths during the 1854 cholera outbreak in Soho, London: a reappraisal of Dr John Snow's investigation. *Health & place, 65*, Article#102402.

Wallert, J., Mattia, T., Guy, M., & Claes, H. (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, no. 1*, Article#99.

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.

Wan Ahmad., W. (2022). *Annual Report of the NCVD-ACS Registry, 2018–2019. Kuala Lumpur, Malaysia.* National Heart Association of Malaysia (NHAM).

Wang, H., Zu, Q., Chen, J., Yang, Z., & Ahmed, M. A. (2021). Application of artificial intelligence in acute coronary syndrome: a brief literature review. *Advances in Therapy*, 1-9.

Wang, L., Liu, C., Meng, X., Niu, Y., Lin, Z., Liu, Y., & Kan, H. (2018). Associations between short-term exposure to ambient sulfur dioxide and increased cause-specific mortality in 272 Chinese cities. *Environment international, 117*, 33-39.

Wang, L., Zhang, Z., Zhang, X., Zhou, X., Wang, P., & Zheng, Y. (2021). A Deep-forest based approach for detecting fraudulent online transaction. In R. H. Ali, & W. Sheng, *Advances in computers* (pp. 1-38). Elsevier.

Wang, X., Kindzierski, W., & Kaul, P. (2015). Air pollution and acute myocardial infarction hospital admission in Alberta, Canada: a three-step procedure case-crossover study. *PloS one, 10(7)*, e0132769.

Wang, Z., Zhang, L., Huang, T., Yang, R., Cheng, H., Wang, H., . . . Lyu, J. (2023). Developing an explainable machine learning model to predict the mechanical ventilation duration of patients with ARDS in intensive care units. *Heart & Lung, 58*, 74-81.

WAQ, W. I. (2020). World's Air Pollution: Real-Time Air Quality Index. Beijing, Beijing, China.

Weber, L., Lapuschkin, S., Binder, A., & Samek, W. (2022). Beyond explaining: Opportunities and challenges of XAI-based model improvement. *Information Fusion*, 154-176.

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.

Westerlund, A., Hawe, J., Heinig, M., & Schunkert, H. (2021). Risk prediction of cardiovascular events by exploration of molecular data with explainable artificial intelligence. *International Journal of Molecular Sciences, 22(19)*, Article#10291.

White, J. E., Wayland, R. A., Dye, T. S., & Chan, A. C. (2004). AIRNow air quality notification and forecasting system. *Beijing International Environment Forum* (pp. 14-15). Beijing: U.S. Environmental Protection Agency.

WHO, W. H. (2018). Ambient air pollution: A global assessment of exposure and burden of disease. *World Health Organization}*, Article#121.

WHO, W. H. (2022). *Air pollution*. Retrieved from World Health Organization: https://www.who.int/health-topics/air-pollution#tab=tab_1

WHO, W. H. (2022). *Noncommunicable diseases.* World Health Organization.

Wongkar, M., & Angdresey, A. (2019). Sentiment analysis using Naive Bayes Algorithm of the data crawler: Twitter. *2019 Fourth International Conference on Informatics and Computing (ICIC)* (pp. 1-5). IEEE.

Wu, T. T., Lin, X. Q., Mu, Y., Li, H., & Guo, Y. S. (2021). Machine learning for early prediction of in-hospital cardiac arrest in patients with acute coronary syndromes. *Clinical Cardiology, 44(3)* , 349-356.

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., . . . Zhou, Z. (2008). Top 10 algorithms in data mining. *Knowledge and information systems, 14*, 1-37.

Wu, Y., Li, R., Cui, L., Meng, Y., Cheng, H., & Fu, H. (2020). The high-resolution estimation of sulfur dioxide (SO2) concentration, health effect and monetary costs in Beijing. *Chemosphere, 241*, Article#125031.

Xiao, Z. (2021). COVID 19 mortality rate prediction based on machine learning methods. *2021 IEEE international conference on computer science, electronic information engineering and intelligent control technology (CEI)* (pp. 169-177). IEEE.

Xu, B., Hui, L., Long S, C., Sammy, T., Jimmy, C., Ya, H., & Liping, Z. (2010). VGE-CUGrid: An integrated platform for efficient configuration, computation, and visualization of MM5. *Environmental Modelling & Software 25, no 12*, 1894-1896.

Xu, S. (2018). Bayesian Naïve Bayes classifiers to text classification. *Journal of Information Science, 44(1)*, 48-59.

Yagin, F. H., Cicek, İ. B., Alkhateeb, A., Yagin, B., Colak, C., Azzeh, M., & Akbulut, S. (2023). Explainable artificial intelligence model for identifying COVID-19 gene biomarkers. *Computers in Biology and Medicine, 154*, Article#106619.

Yamamoto, S. S., Phalkey, R., & Malik, A. A. (2014). A systematic review of air pollution as a risk factor for cardiovascular disease in South Asia: Limited evidence from India and Pakistan. . *International journal of hygiene and environmental health, 217(2-3)*, 133-144.

Yang, B., Guo, Y., Markevych, I., Qian, Z., Bloom, M., Heinrich, J., . . . Leskinen, A. (2019). Association of long-term exposure to ambient air pollutants with risk factors for cardiovascular disease in China. *JAMA network open, 2(3)*, e190318-e190318.

Yao, M., Wu, G., Zhao, X., & Zhang, J. (2020). Estimating health burden and economic loss attributable to short-term exposure to multiple air pollutants in China. *Environmental Research, 183*, Article#109184.

Yatsuya, H. (2018). Risk Prediction. In S. V. Ramachandran, & B. S. Douglas, *Encyclopedia of Cardiovascular Research and Medicine* (pp. 315-318). Elsevier.

Yuan, S., Wang, J., Jiang, Q., He, Z., Huang, Y., Li, Z., . . . Cao, S. (2019). Long-term exposure to PM2. 5 and stroke: a systematic review and meta-analysis of cohort studies. *Environmental research, 177*, Article#108587.

Zając, A. (2023, February 01). *TIMI Score for STEMI Calculator*. Retrieved from OMNI Calculator: https://www.omnicalculator.com/health/timi-stemi

Zając, A. (2023b, January 18). *TIMI Score Calculator for UA/NSTEMI*. Retrieved from OMNI Calculator: https://www.omnicalculator.com/health/timi-ua-nstemi

Zambahari, R. (2004). Trends in cardiovascular diseases and risk factors in Malaysia. *Elsevier*, 446-449.

Zanobetti, A., Schwartz, J., Samoli, E., Gryparis, A., Touloumi, G., Peacock, J., . . . Goren, A. (2003). The temporal pattern of respiratory and heart disease mortality in response to air pollution. *Environmental health perspectives, 111(9)*, 1188-1193.

Zeng, Z. (2022). Explainable artificial intelligence (XAI) for healthcare decision-making. *Nanyang Technological University*, 1-18.

Zhalehdoost, A., & Taleai, M. (2022). A Review of the Application of Machine Learning and Geospatial Analysis Methods in Air Pollution Prediction. *Pollution, 8(3)*, 904-933.

Zhang, C., Ding, R., Xiao, C., Xu, Y., Cheng, H., Zhu, F., . . . Cao, J. (2017). Association between air pollution and cardiovascular mortality in Hefei, China: a time-series analysis. *Environmental Pollution, 229*, 790-797.

Zhang, S., Yuan, Y., Yao, Z., Yang, J., Wang, X., & Tian, J. (2022). Coronary Artery Disease Detection Model Based on Class Balancing Methods and LightGBM Algorithm. *Electronics, 11(9)*, Article#1495.

Zhang, X., Fung, J. C., Lau, A. K., Hossain, M. S., Louie, P. K., & Huang, W. (2021). Air quality and synergistic health effects of ozone and nitrogen oxides in response to China's integrated air quality control policies during 2015–2019 . *Chemosphere, 268*, Article#129385.

Zhang, Y., Liu, F., Zhao, Z., Li, D., Zhou, X., & Wang, J. (2012). Studies on application of Support Vector Machine in diagnose of coronary heart disease. *In Sixth International Conference on Electromagnetic Field Problems and Applications* (pp. 1-4). IEEE.

Zhang, Z., Chen, L., Xu, P., & Hong, Y. (2022). Predictive analytics with ensemble modeling in laparoscopic surgery: a technical note. *Laparoscopic, Endoscopic and Robotic Surgery, 5(1)*, 25-34.

Zhao, B., Johnston, F. H., Salimi, F., Oshima, K., Kurabayashi, M., & Negishi, K. (2023). Short-term exposure to sulfur dioxide and nitrogen monoxide and risk of out-of-hospital cardiac arrest. *Heart, Lung and Circulation, 32(1)*, 59-66.

Zhao, B., Salimi, F., Johnston, F., Oshima, K., Kurabayashi, M., & Negishi, K. (2016). Out-of-Hospital Cardiac Arrest and Short-Term Exposure to Air Pollutants: A Case-Crossover Study. *Circulation, 134(suppl_1)*, A16275-A16275.

Zheng, B., Huo, Y., Lee, S., Sawhney, J., Kim, H., Krittayaphong, R., . . . Jiang, J. (2020). Long-term antithrombotic management patterns in Asian patients with acute coronary syndrome: 2-year observations from the EPICOR Asia study. *Clinical cardiology, 43(9)*, 999-1008.

Zheng, H., Sherazi, S., & Lee, J. (2021). A stacking ensemble prediction model for the occurrences of major adverse cardiovascular events in patients with acute coronary syndrome on imbalanced data. *IEEE Access, 9*, 113692-113704.

Zhou, N., Ji, Z., Li, F., Qiao, B., Lin, R., Jiang, W., . . . You, B. (2022). Machine Learning-Based Personalized Risk Prediction Model for Mortality of Patients Undergoing Mitral Valve Surgery: The PRIME Score. *Frontiers in Cardiovascular Medicine*, 9.

Zhou, Y., Wang, H., & Liu, H. (2019). Generalized function projective synchronization of incommensurate fractional-order chaotic systems with inputs saturation. *International Journal of Fuzzy Systems, vol. 21, no. 3*, 823–836.

Zhu, F., Li, X., Tang, H., He, Z., Zhang, C., Hung, G., . . . Zhou, W. (2020). Machine learning for the preliminary diagnosis of dementia. *Scientific Programming, 2020*, 1-10 .

Zisou, C., Sochopoulos, A., & Kitsios, K. (2020). Convolutional recurrent neural network and LightGBM ensemble model for 12-lead ECG classification. *2020 Computing in Cardiology* (pp. 1-4). IEEE.