

**Development of a Cardiovascular Disease Risk Prediction System using Machine Learning and Patient Health Data**

**\_\_\_\_\_\_\_\_\_\_**

**Jason Jay Dookarun**

**URN: 6494278**

**Supervisor: Dr Tom Thorne**

**Department of Computer Science**

**University of Surrey**

Declaration

I, Jason Jay Dookarun, of the Department of Computer Science, University of Surrey, confirm that this is my own work and figures, tables, equations, code snippets, artworks, and illustrations in this report are original and have not been taken from any other person's work, except where the works of others have been explicitly acknowledged, quoted, and referenced. I understand that if failing to do so will be considered a case of plagiarism. Plagiarism is a form of academic misconduct and will be penalised accordingly.

**Jason Jay Dookarun**

**May 23rd, 2023**

Abstract

Cardiovascular diseases (CVDs) are a significant global health concern, demanding the development of effective risk prediction systems for early identification and intervention. This thesis focuses on the development of a cardiovascular disease risk prediction system using machine learning algorithms and patient health data.

The study begins with a comprehensive review of relevant literature, exploring existing risk assessment models and identifying gaps in current approaches. A thorough analysis of available patient health data is conducted, encompassing clinical measurements, demographic information, and lifestyle indicators, to capture a comprehensive understanding of the factors contributing to cardiovascular disease risk.

Various statistical models, including logistic regression, decision trees, random forests, and support vector machines, are examined and compared for their predictive performance. The models are trained and evaluated using a large-scale dataset containing anonymized patient records, allowing for the identification of patterns and risk factors associated with cardiovascular diseases.

The outcome of this research is a robust cardiovascular disease risk prediction system that enables healthcare professionals to assess an individual's propensity for cardiovascular disease based on their health data. The developed system provides a personalized risk score for each patient, empowering clinicians to make informed decisions regarding preventive interventions and patient management strategies.

The evaluation of the developed system includes performance metrics such as accuracy, precision and recall, ensuring its reliability and effectiveness. Additionally, potential limitations and areas for future research are discussed to guide further advancements in cardiovascular disease risk prediction using statistical modelling techniques.

Overall, this thesis contributes to the field of cardiovascular disease prevention by integrating statistical models and patient health data to develop an accurate and personalized risk prediction system. The findings have significant implications for early intervention and tailored healthcare strategies, ultimately reducing the burden of cardiovascular diseases on individuals and healthcare systems.

Acknowledgements

Table of Contents

[Declaration 2](#_Toc135764013)

[Abstract 3](#_Toc135764014)

[Acknowledgements 4](#_Toc135764015)

[Introduction 6](#_Toc135764016)

[Literature Review 6](#_Toc135764017)

[Methodology 7](#_Toc135764018)

[Results 8](#_Toc135764019)

[Discussion 9](#_Toc135764020)

[Conclusion & Future Work 10](#_Toc135764021)

[Reflections 11](#_Toc135764022)

[References 12](#_Toc135764023)

[Appendices 13](#_Toc135764024)

Introduction

Cardiovascular diseases also known as CVDs have emerged to be one of the greatest contributors to causes of death over the past decades, accounting to a substantial proportion of mortality worldwide. Around 66,000 deaths in the UK, an average of 180 people die from coronary heart disease (British Heart Foundation, 2023). Coronary heart diseases range from a multitude of conditions that affect the heart. These can range from conditions that affect heart and blood vessels, including coronary artery disease, myocardial infraction, also known as heart attacks and strokes. Similarly, CVDs represent additional diseases such as cerebrovascular disease, rheumatic heart disease and other conditions. According to the World Health Organisation, cardiovascular diseases (CVDs) are the “leading cause of death globally, responsible for approximately 17.9 million deaths each year” (World Health Organisation , 2021), with this statistic being published pre-COVID 2019 pandemic.

Age, genetic predisposition, lifestyle choices, and underlying health conditions all influence the prevalence of cardiovascular diseases. Smoking, physical inactivity, poor diet, and excessive alcohol consumption all contribute to the development of risk factors such as hypertension, obesity, high cholesterol levels, and diabetes. These risk factors, in turn, raise the chances of developing cardiovascular disease.

Literature Review

Methodology

Results

Discussion

Conclusion & Future Work

Reflections

References

Appendices