****

M.Sc. Data Analytics and Technologies

**University of Bolton**

School of Creative Technologies

**Development of Advanced Ensembling Machine Learning algorithms for improved mortality prediction in open-heart surgery patients**

Module leader:

**Dr Celestin Iwendi**

Module Name:

**SWE7004 - Research Methods**

**Fateme Mousavi Moayed (2218697)**

May 2023

Table of Contents

[Table of Tables 3](#_Toc135328230)

[Table of Abbreviations 4](#_Toc135328231)

[Abstract 5](#_Toc135328232)

[Introduction 6](#_Toc135328233)

[Background and Context 6](#_Toc135328234)

[Problem Statement 6](#_Toc135328235)

[Research Questions 7](#_Toc135328236)

[Research Aims and Objectives 7](#_Toc135328237)

[Overall Aim 8](#_Toc135328238)

[Specific Objectives 8](#_Toc135328239)

[Relevance and Importance of the Research 8](#_Toc135328240)

[Literature Review 11](#_Toc135328241)

[Different Machine Learning Models 11](#_Toc135328242)

[Importance of Machine Learning Algorithms 12](#_Toc135328243)

[Ensemble Methods 12](#_Toc135328244)

[Optimization Techniques 13](#_Toc135328245)

[Literature on Topic and Method 15](#_Toc135328246)

[Literature on Algorithms 19](#_Toc135328247)

[Research Design and Methods 25](#_Toc135328248)

[Research Methodology 25](#_Toc135328249)

[Quantitative Approach 25](#_Toc135328250)

[Qualitative Approach 27](#_Toc135328251)

[Implications and contributions to knowledge 29](#_Toc135328252)

[References 31](#_Toc135328253)

# Table of Tables

Table 1: Summary Table of related Academic Papers………………………………………………7

Table 2: Timetable for this research……………………………………………………………….………13

**Table of Figures**

# Table of Abbreviations

# Abstract

# Introduction

Despite open-heart surgery has saved lots of lives, predicting mortality risk for patients remains a significant challenge. Machine Learning (ML) techniques have positive performance in risk of mortality, but the accuracy of these predictions depends on the algorithm and data used (Ahdal et al., 2021; Li et al., 2020).

## Background and Context

In recent years, the field of ML has emerged as a powerful tool for analysing vast amounts of clinical data and extracting valuable insights to aid in medical decision-making (Ahmad et al., 2022; Ramesh et al., 2022). Traditional ML algorithms have been widely explored for mortality prediction in various healthcare domains (Ashri et al., 2021). However, due to the complexity and heterogeneity of open-heart surgery patients, these algorithms often struggle to achieve satisfactory predictive accuracy (Sarra et al., 2022). To overcome these limitations, researchers have turned to advanced assembling ML techniques that combine multiple algorithms to enhance predictive performance. Assembling ML algorithms, such as ensemble methods and stacking models, has shown promise in various healthcare applications by leveraging the strengths of multiple individual models (Zhang et al., 2021).

## Problem Statement

There is still a knowledge gap about how to increase predictive accuracy, despite earlier studies using ML approaches to estimate mortality risk in patients undergoing open-heart surgery (Li et al., 2020; Ramesh et al., 2022). Some ML algorithms have shown some results, but they have limitations in accurately predicting mortality risk due to the complexity and heterogeneity of these patients (Ashri et al., 2021; Sarra et al., 2022). Advanced assembling ML algorithms, such as ensemble methods and stacking models, have shown promise in healthcare applications by combining multiple models to improve predictive accuracy. Therefore, the main problem is finding ways to improve the accuracy of predicting mortality risk in open-heart surgery patients by developing advanced ensembling ML algorithms.

## Research Questions

The main objective of this study is to explore the use of advanced ensembling ML algorithms for improved mortality prediction accuracy in open-heart surgery patients. The specific research questions include:

* What are the current limitations and uses of ML algorithms for mortality prediction in open-heart surgery patients?
* How can ensembling improve predictive accuracy?
* What is the importance of data features for predictive accuracy?
* What are the potential clinical implications of improved prediction?

# Research Aims and Objectives

The successful development of advanced assembling ML algorithms for improved mortality prediction in open-heart surgery patients holds significant potential for enhancing clinical decision-making and patient care. The purpose of this research is to explore whether assembling ML algorithms boost accuracy in patients undergoing open-heart surgery. In this regard, this study aims to undertake the following objectives.

## Overall Aim

The general aim of this study is to find the potential of ensembling ML algorithms to enhance the accuracy of mortality prediction in patients undergoing open-heart surgery.

## Specific Objectives

1. Review current ML algorithms for mortality prediction in open-heart surgery.
2. Identify limitations of current ML algorithms for mortality prediction in open-heart surgery.
3. Evaluate ensembling ML models to improve mortality prediction in open-heart surgery.
4. Identify important features for mortality prediction in open-heart surgery.
5. Assess the clinical implications of improved mortality prediction in open-heart surgery.

# Relevance and Importance of the Research

The research on developing advanced assembling ML algorithms for improved mortality prediction in open-heart surgery patients is highly relevant to cardiovascular medicine. By identifying high-risk patients, clinicians can implement appropriate strategies to optimize perioperative care, potentially reducing adverse outcomes and improving patient survival rates. It is critical to boost patient outcomes and informed clinical decision-making to Improve the accuracy of these prediction models. This study highlighted the gaps in existing knowledge and explored the potential of ensembling techniques to enhance the accuracy of mortality prediction models for open-heart surgery patients. The results of this study help researchers in this field to take better care of the patient and improved medical practices.

# Literature Review

## Different Machine Learning Models

In the field of mortality prediction for patients undergoing open-heart surgery, various ML models have been applied to analyse complex clinical data and make accurate predictions. These models leverage the power of computational algorithms to uncover patterns and relationships within the data, enabling healthcare professionals to assess mortality risk more effectively.

One commonly used ML model is the Decision Tree algorithm. Hierarchical structures known as decision trees utilize a series of binary decisions to make predictions. The interpretation of these models is straightforward, and they offer valuable insights into the decision-making process. Decision Trees have been employed in mortality prediction to identify relevant features and classify patients into different risk categories.

Another popular model is the Random Forest algorithm, which combines multiple Decision Trees to improve prediction accuracy. Random Forest leverages the concept of ensemble learning, where predictions from multiple models are aggregated to make a final prediction. By combining diverse Decision Trees, Random Forest reduces overfitting and captures complex relationships in the data, leading to more accurate mortality predictions.

Support Vector Machines (SVMs) have also been widely utilized in mortality prediction. SVMs aim to find an optimal hyperplane that separates different classes in the data by maximizing the margin between them. These models can handle high-dimensional data and are effective in capturing non-linear relationships. SVMs have been applied to identify patients at risk for specific complications following open-heart surgery.

## Importance of Machine Learning Algorithms

Traditional statistical methods may struggle to handle the complexity and non-linearity of clinical data, whereas ML algorithms excel in analysing large and intricate datasets. These algorithms can identify hidden patterns, interactions, and predictive relationships that may not be apparent through conventional statistical approaches. By accurately predicting mortality risk, healthcare providers can proactively identify patients who require closer monitoring or specific interventions, such as post-operative care protocols or personalized treatment plans. The importance of ML algorithms lies in their ability to support evidence-based decision-making and enhance clinical practice in open-heart surgery.

## Ensemble Methods

Ensemble methods combine multiple ML models to make more accurate predictions. They leverage the wisdom of crowds, where the collective decision of multiple models tends to be more reliable than that of individual models. Ensemble methods have been successfully applied in mortality prediction for open-heart surgery patients to improve prediction accuracy and robustness. One popular ensemble method is Bagging, which generates multiple models by training them on different subsets of the data using bootstrapping. The predictions from each model are then combined, often by averaging or voting, to make the final prediction. Bagging reduces the variance of individual models and can mitigate the impact of outliers or noise in the data. Another ensemble method is Boosting, which iteratively builds models that focus on the misclassified instances from previous models. Boosting assigns higher weights to these instances, effectively prioritizing the challenging cases. This iterative process leads to a strong ensemble model that can accurately capture complex patterns and improve mortality prediction performance.

**Feature Selection Algorithms**

Feature selection algorithms aim to identify the most informative and relevant features in the dataset, enabling accurate predictions. By selecting the key risk factors, feature selection plays a critical role in understanding the contributors to mortality outcomes.

One commonly used feature selection algorithm is Recursive Feature Elimination (RFE). RFE works by recursively eliminating features from the dataset and assessing the impact on model performance. It ranks the features based on their importance and selects the optimal subset of features that maximizes prediction accuracy.

Another popular feature selection algorithm is the Minimum Redundancy Maximum Relevance (mRMR) method. mRMR measures the relevance and redundancy of features, aiming to select a subset that contains highly relevant features while minimizing redundancy. By selecting a concise yet informative set of features, mRMR enhances the interpretability and performance of mortality prediction models.

## Optimization Techniques

Optimization techniques are used to fine-tune ML models and improve their performance. These techniques search for the optimal set of model parameters or hyperparameters that maximize prediction accuracy.

GridSearchCV is a commonly used optimization technique that systematically explores the hyperparameter space of ML models. It evaluates the performance of models across different hyperparameter combinations and identifies the best set of parameters that yield the highest accuracy. GridSearchCV automates the process of parameter tuning and helps researchers find the optimal configuration for mortality prediction models.

Genetic algorithms provide another optimization approach. Inspired by the process of natural selection, genetic algorithms iteratively evolve a population of potential solutions. They simulate genetic operations such as mutation and crossover to produce new candidate solutions. Genetic algorithms have been applied in mortality prediction to optimize feature selection, hyperparameter tuning, or model architecture. These optimization techniques enable researchers to maximize the performance of mortality prediction models by finding the optimal combination of parameters or hyperparameters. By fine-tuning the models, researchers can achieve higher accuracy and improve the reliability of mortality predictions for open-heart surgery patients.

In summary, the utilization of different ML models, the importance of ML algorithms, the application of ensemble methods, the use of feature selection algorithms, and the employment of optimization techniques collectively contribute to enhanced mortality prediction accuracy in patients undergoing open-heart surgery. These techniques empower healthcare providers with reliable tools for risk assessment, personalized interventions, and improved patient outcomes. By leveraging the power of ML algorithms and optimizing their performance, researchers strive to improve mortality prediction and ultimately enhance healthcare delivery in the field of open-heart surgery. All these concepts and phases are shown in Figure 1.

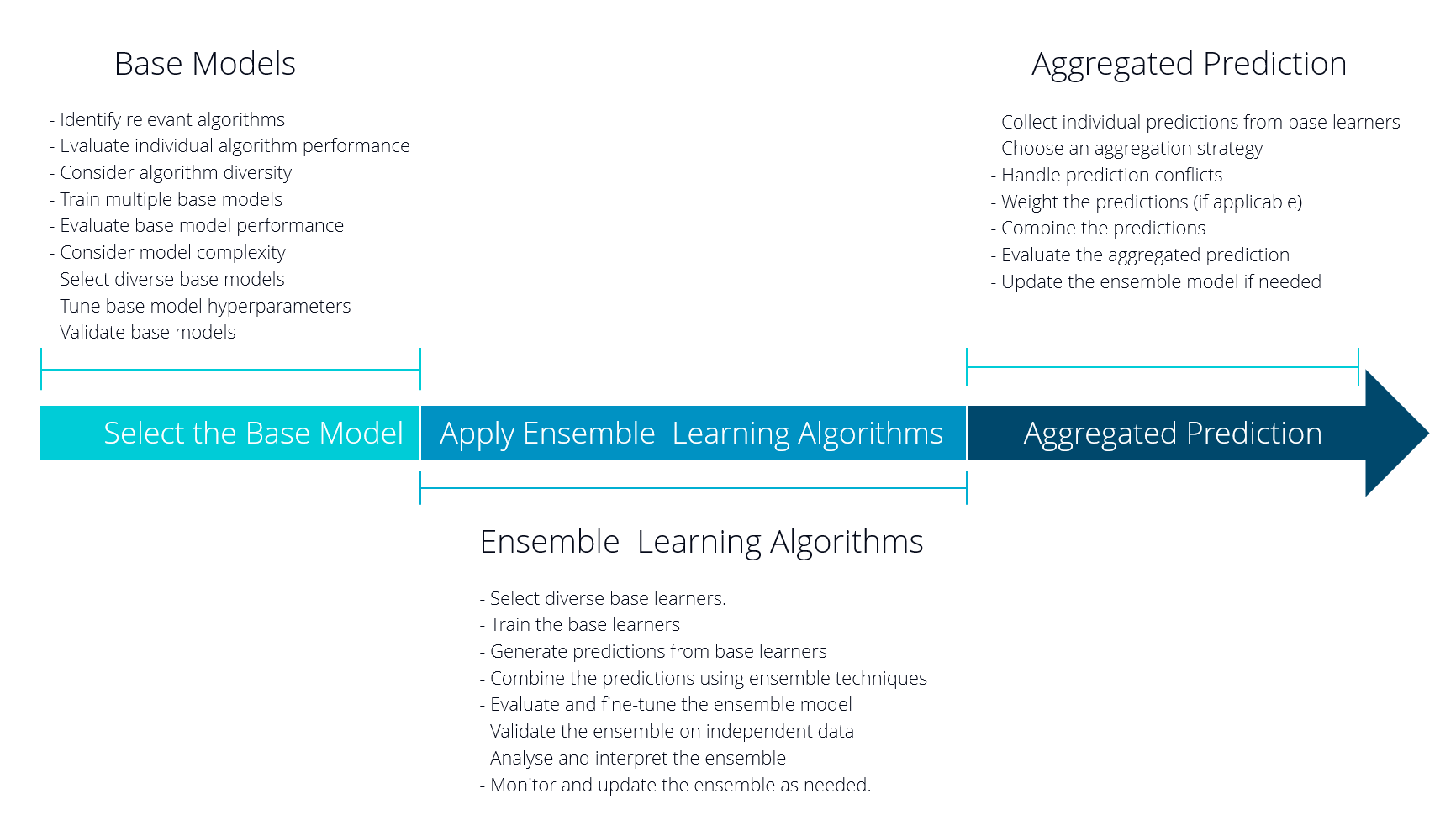


Figure 1. Concepts and phase of the research

## Literature on Topic and Method

Mortality prediction in open-heart surgery patients has seen advancements through ML techniques and algorithm exploration. ML techniques have performed considerable attention in this domain due to their ability to analyse complex clinical data and identify predictive patterns. This review examines the literature on the topic and method, and algorithms used, providing insights for improved mortality prediction in open-heart surgery.

Ahdal et al. (2021) discussed the use of ML for heart patient scanning, visualization, and monitoring. They highlighted the potential of these techniques to enhance diagnosis and risk assessment in cardiac patients. By employing ML algorithms, they were able to detect abnormal patterns in cardiac data, leading to the early detection of potential issues and improving patient care. this study highlighted the significance of leveraging ML techniques to improve patient monitoring and intervention in the context of open-heart surgery.

Ahmad et al. (2022) proposed an efficient medical diagnosis of heart diseases using ML techniques, including GridSearchCV. They emphasized the importance of accurate diagnosis for appropriate treatment planning. Their study demonstrated the effectiveness of ML algorithms in accurately classifying different types of heart diseases, enabling healthcare providers to make informed decisions regarding patient treatment plans. By utilizing GridSearchCV, which systematically explores the hyperparameter space of ML algorithms, they were able to optimize the model's performance and improve diagnostic accuracy.

Ashri et al. (2021) developed the Heart Disease Prediction Framework (HDPF) based on hybrid classifiers and genetic algorithms for the proper prediction of heart diseases. The hybrid classifiers merged the strengths of multiple algorithms, namely K-Nearest Neighbours (KNN), Decision Tree, and Naive Bayes, to improve the prediction accuracy. the genetic algorithm, Additionally, was employed to optimize the feature selection process, enhancing the overall performance of the prediction model. The study highlighted the importance of incorporating multiple algorithms and feature selection techniques to achieve accurate mortality prediction in open-heart surgery patients.

Bertsimas et al. (2021) focused on real-time heart disease prediction using ML. They emphasized the need for timely interventions and the potential of ML algorithms to enable the early detection of heart diseases. The study explored various ML techniques, including Support Vector Machines (SVM), Random Forest, and Gradient Boosting, and demonstrated their effectiveness in predicting heart disease outcomes. The authors also highlighted the importance of interpretable models in the healthcare domain to gain insights into the factors contributing to mortality prediction.

Fitriyani et al. (2020) proposed HDPM, an effective heart disease prediction model for a clinical decision support system. They operated ML algorithms namely K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest (RF). The study showcased the advantages of ML in accurately predicting heart disease outcomes and the potential of implementing such models in clinical decision support systems. The authors emphasized the importance of incorporating different algorithms to capture diverse patterns and improve prediction accuracy.

Garg et al. (2021) explored the use of ML techniques, namely Naive Bayes, Random Forest, and Support Vector Machine, for heart disease prediction. Their study focused on the application of ML algorithms to predict heart disease and highlighted the potential of these techniques to identify individuals at risk. The authors discussed the importance of feature selection and model evaluation techniques to improve the performance of the prediction models.

Gupta et al. (2020) proposed MIFH, a machine intelligence framework for heart disease diagnosis. The framework utilized ML algorithms such as Decision Trees, Random Forests, and K-Nearest Neighbours for accurate diagnosis. The authors highlighted the significance of interpretable models in the healthcare domain and their potential to assist healthcare providers in making informed decisions regarding treatment and management strategies for open-heart surgery patients. By utilizing interpretable models such as Decision Trees, healthcare providers can gain insights into the decision-making process of the model and understand the contributing factors leading to a particular diagnosis. This transparency and interpretability can enhance the trust and acceptance of ML algorithms in clinical settings.

He et al. (2022) developed a model based on photoplethysmography and ML to predict haemorrhagic risk in coronary artery disease patients. They utilized the Gradient Boosting Machine (GBM) algorithm to achieve accurate predictions. Photoplethysmography, a non-invasive optical technique, measures changes in blood volume, providing valuable information about the patient's vascular condition. By combining photoplethysmography data with ML algorithms, the model identifies patients at a higher risk of haemorrhage and provides early intervention, potentially improving patient outcomes. This study demonstrated the potential of incorporating physiological signals and advanced ML algorithms in mortality prediction for open-heart surgery patients.

Huda et al. (2021) proposed an ML model for identifying patients at risk for wild-type transthyretin amyloid cardiomyopathy. They employed ML algorithms such as Support Vector Machines (SVM), Random Forest, and Gradient Boosting to develop the prediction model. The study concentrated on determining specific biomarkers and clinical features related to the development of amyloid cardiomyopathy, a progressive heart disease. By leveraging ML techniques, the model could accurately identify individuals at risk, enabling early intervention and improved patient management strategies.

Niimi et al. (2022) investigated ML models for the prediction of adverse events after percutaneous coronary intervention. They utilized algorithms such as Random Forest and Deep Neural Networks to improve risk assessment. By analysing various clinical and procedural factors, the ML models could predict adverse events with higher accuracy compared to traditional risk scoring systems. This study highlights the potential of ML in identifying patients who may be at a higher risk of complications following percutaneous coronary intervention, allowing healthcare providers to take necessary precautions and provide appropriate post-procedural care.

Qian et al. (2022) developed a cardiovascular disease prediction model based on routine physical examination indicators using ML methods. Their study focused on leveraging easily accessible clinical data to predict cardiovascular disease outcomes. The ML algorithms employed, including Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine, demonstrated the ability to accurately predict cardiovascular disease risk based on routine physical examination indicators. This research showcases the potential of ML techniques to facilitate early detection and risk assessment of cardiovascular diseases, aiding in the prevention and management of heart conditions.

In summary, the literature on the topic and methods highlight the significance of ML techniques, such as Decision Trees, Random Forests, Support Vector Machine, and Gradient Boosting, in mortality prediction for open-heart surgery patients. These algorithms, combined with interpretable models, advanced physiological signals, and clinical data, have the potential to improve risk assessment, diagnosis, and treatment planning for patients undergoing open-heart surgery. By leveraging ML techniques, healthcare providers make more informed decisions, personalize interventions, and ultimately improve patient outcomes.

## Literature on Algorithms

ML algorithms play a crucial role in mortality prediction for open-heart surgery patients. Various algorithms have been explored to analyse clinical data, identify patterns, and make accurate predictions.

One commonly employed algorithm is the Decision Tree algorithm. Penny-Dimri et al. (2021) utilized Decision Trees to predict and risk profile cardiac surgery-associated acute kidney injury. Decision Trees provide a transparent and interpretable model structure by splitting the data based on different features, leading to the creation of a tree-like model. The authors highlighted the significance of Decision Trees in predicting acute kidney injury following cardiac surgery, which enables early intervention and appropriate management strategies.

Random Forest, an ensemble learning algorithm, has been widely adopted in mortality prediction for open-heart surgery patients. Medved et al. (2018) improved the prediction of heart transplantation outcomes using Random Forest and deep learning techniques. By combining multiple decision trees, Random Forest captures diverse patterns and reduces the risk of overfitting, leading to improved prediction accuracy. The authors demonstrated that Random Forest outperformed traditional models, providing more accurate risk stratification for heart transplant recipients.

Gradient Boosting algorithms, such as XGBoost and LightGBM, have shown promise in mortality prediction for open-heart surgery patients. Ahmad et al. (2018) utilized Gradient Boosting algorithms to improve prognostication and identify clinically distinct phenotypes in heart failure patients. These algorithms iteratively optimize the model by combining weak classifiers, building a strong predictive model. The authors emphasized that Gradient Boosting algorithms can capture complex relationships and heterogeneity in response to therapy, enabling more accurate risk stratification and personalized treatment decisions.

Support Vector Machine (SVM) is another popular algorithm for open-heart surgery patients. Cheng et al. (2021) employed SVM to indicate hospital-acquired thrombocytopenia after surgery in the intensive care unit. SVM aims to discover the optimal hyperplane that separates different classes in the data. By employing SVM, the study accurately predicts the occurrence of thrombocytopenia, facilitating early identification and intervention for patients at risk.

Another algorithm used in mortality prediction is K-Nearest Neighbours (KNN). Huda et al. (2021) employed KNN to identify patients at risk for wild-type transthyretin amyloid cardiomyopathy. KNN is a non-parametric algorithm that classifies data points based on their proximity to neighbouring points. The authors demonstrated that KNN, by considering the nearest neighbours of patients, could accurately identify individuals at risk of developing amyloid cardiomyopathy, enabling early detection and intervention.

DL algorithms, namely Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), have also been applied in mortality prediction for open-heart surgery patients. Deo (2015) discussed the use of DL in medicine, emphasizing its potential to extract meaningful features and patterns from complex medical data. These algorithms, with their ability to learn hierarchical representations and capture temporal dependencies, have the potential to improve mortality prediction accuracy by effectively leveraging large-scale clinical data.

In addition to these algorithms, ensemble methods have gained attention for mortality prediction in open-heart surgery patients. Jindal et al. (2021) explored the performance of ensemble algorithms, namely Random Forest, Gradient Boosting, and Extreme Gradient Boosting, in heart disease prediction. Ensemble methods combine the predictions of multiple individual models to produce a final prediction. By aggregating diverse predictions, ensemble methods can improve prediction accuracy and robustness.

Feature selection algorithms are also crucial in mortality prediction. Wang et al. (2022) employed the minimum Redundancy Maximum Relevance (mRMR) method for heart disease diagnosis. This algorithm aims to select a subset of features that are both informative and non-redundant. By utilizing the mRMR method, the authors were able to identify the most relevant features for accurate diagnosis, reducing the dimensionality of the data and improving the performance of mortality prediction models.

In addition to traditional ML algorithms, advanced optimization techniques have been applied to mortality prediction in open-heart surgery patients. Ashri et al. (2021) utilized a genetic algorithm to optimize the feature selection process in their Heart Disease Prediction Framework (HDPF). Genetic algorithms are evolutionary search algorithms that mimic the process of natural selection. By iteratively evolving a population of potential solutions, genetic algorithms can identify the optimal combination of features for improved prediction accuracy.

Furthermore, ensemble learning techniques have been employed to enhance the predictive performance of ML models in mortality prediction. Ahmad et al. (2022) utilized GridSearchCV, a technique for hyperparameter tuning, in their efficient medical diagnosis of heart diseases. GridSearchCV systematically explores the hyperparameter space of ML algorithms to identify the optimal set of hyperparameters that yield the best performance. By employing GridSearchCV, the authors were able to fine-tune the parameters of their models, improving the accuracy of heart disease diagnosis.

Moreover, DL techniques have shown promise in mortality prediction for open-heart surgery patients. Medved et al. (2018) utilized deep learning techniques, specifically Convolutional Neural Networks (CNNs), to improve the prediction of heart transplantation outcomes. CNNs excel at learning intricate spatial patterns in data, making them suitable for analysing complex medical images or time-series data. By leveraging the hierarchical representations learned by CNNs, the authors were able to extract meaningful features from medical images, enhancing the accuracy of transplantation outcome prediction. Selecting the algorithms depends on various aspects, namely the nature of the data, the complexity of the problem, and the available computational resources. Researchers explore and develop algorithms and also optimise techniques to boost the result of accuracy and efficiency.

The literature on algorithms for mortality prediction in open-heart surgery patients highlights the effectiveness of various ML Algorithms, including Decision Trees, Random Forest, Gradient Boosting, Support Vector Machines, K-Nearest Neighbours, and DL techniques namely CNNs. Ensemble methods, feature selection algorithms, and optimization techniques such as genetic algorithms and GridSearchCV have also been employed to enhance the predictive performance of models. Each algorithm has strengths and could be applied based on the specific characteristics of the dataset and the objectives of the prediction task. By leveraging these algorithms, researchers aim to improve mortality prediction accuracy and provide a valuable understanding of clinical decision-making.

Table 1: Summary Table of related Academic Papers

Table

Description automatically generated with medium confidence

# Research Design and Methods

## Research Design

### Quantitative Approach

The research design will use a quantitative methodology to investigate the efficacy of combining ML algorithms in predicting mortality for patients undergoing open-heart surgery. This study uses secondary data sources and is descriptive and correlational in nature. In Figure 2, steps of this research are shown.

A diagram of a model development process

Description automatically generated with low confidence

Figure 2. Steps of the research (Machine Learning to ensemble)

The research will follow the CRISP-DM (Cross Industry Standard Process for Data Mining) approach to develop advanced ensembling ML algorithms for improved mortality prediction in open-heart surgery patients. As shown in Figure 3 below, the CRISP-DM approach consists of six phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment (Chapman et al., 2000).

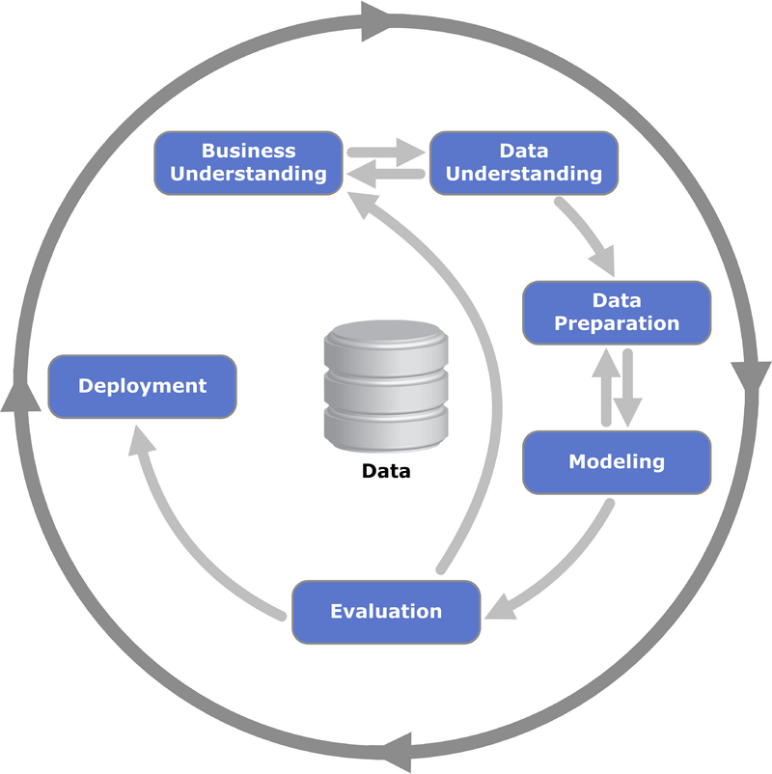


Figure 3. The CRISP-DM methodology (Chapman et al., 2000)

The research will begin with the Business Understanding phase, where the research question will be defined, and the objectives of the study will be established. In the Data Understanding phase, publicly available datasets will be collected, and a thorough review of essential findings from ML algorithm studies on open-heart surgery mortality prediction will be conducted. The Data Preparation phase will involve data cleaning, integration, and transformation. The Modelling phase will develop advanced ensembling ML algorithms to predict mortality for patients undergoing open-heart surgery. The Evaluation phase will evaluate the performance of the model’s using accuracy, sensitivity, and specificity. Finally, the Deployment phase will involve the implementation and deployment of the models in clinical practice. This study will use ML algorithms such as decision trees, random forests, and support vector machines to analyse the data. Individual and ensemble models' performance will be evaluated based on their accuracy, sensitivity, and specificity. The relevance of data features will be evaluated using feature selection techniques, such as recursive feature elimination.

### Qualitative Approach

The research design will also employ a qualitative methodology to conduct interviews with healthcare professionals involved in open-heart surgery patient care. These interviews aim to capture the insights, experiences, and perspectives of healthcare professionals regarding mortality prediction and the potential use of ensembling ML algorithms in their clinical practice. By engaging with healthcare professionals, the research seeks to understand their perceptions, challenges, and expectations related to mortality prediction, as well as their views on the integration of advanced ML techniques. The Qualitative data from these interviews provide valuable insights into the practical considerations, ethical concerns, and potential benefits of ensembling algorithms to predict mortality after open-heart surgery. Incorporating these perspectives into the research findings contribute to a comprehensive understanding of the topic and informs the development of evidence-based recommendations for improving patient outcomes and clinical decision-making in the field of cardiac medicine.

The research considers and addresses possible obstacles and limitations, including data availability, quality, and algorithmic performance. The study addresses the potential issue of overfitting by conducting cross-validation experiments to ensure that the models are not just learning the noise in the data. The ethical considerations of data protection, informed consent, and confidentiality also is considered. The University RE1 ethics form is attached as Appendix B.

In Figure 4, both Quantitative and Qualitative Approach are shown.

A picture containing text, screenshot, font, design

Description automatically generated

Figure 4. Research Methodologies (Quantitative and Qualitative Approach)

## Methods and Sources

In this Research feature selection techniques, cross-validation experiments and ML algorithms (like decision trees, random forests, and support vector machines) are used. The research follows the CRISP-DM approach, including phases such as Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment.

The study relies on secondary data sources, and no specific participants are involved. Publicly available datasets and existing literature on ML algorithm studies related to open-heart surgery mortality prediction. Additionally, the research design incorporates a qualitative approach with interviews conducted among healthcare professionals involved in open-heart surgery patient care. These interviews provide insights into practical considerations, ethical concerns, and potential benefits of ensembling ML algorithms in clinical practice.

## Practical Consideration

To ensure a smooth execution of the study, potential obstacles, limitations, and practical issues will be carefully addressed. Data availability and quality will be thoroughly assessed to ensure the reliability of the findings. Additionally, algorithmic performance will be evaluated and optimized to enhance the accuracy of mortality predictions. Cross-validation experiments will be conducted to mitigate the risk of overfitting and ensure the models generalize well to unseen data.

## Ethical Consideration

Ethical considerations are of utmost importance in this study, and all necessary measures will be taken to ensure the ethical integrity of the research. The University RE1 ethics form, attached as Appendix A, outlines the specific protocols and procedures that will be followed to protect the rights and well-being of the participants, ensure data protection, and maintain confidentiality throughout the study. Informed consent will be obtained from participants involved in data collection or interviews, and strict data privacy protocols will be implemented. Additionally, **Figure 5 and Figure 6,** present the agreement of the application for data access, where the researcher applied for access to the database on April 13, 2023, and the application was accepted by The PhysioNet Team (MIT Laboratory for Computational Physiology Institute for Medical Engineering and Science) on the same date. These steps demonstrate the researcher's commitment to ethical research practices A picture containing text, screenshot, font, letter

Description automatically generatedand compliance with the University and PhysioNet guidelines.

Figure 5. The confirmation of the application for data access request

A screenshot of a computer

Description automatically generated

Figure 6. The agreement of the application for data access

# Implications and Contributions to Knowledge

## Practical implications

By improving the accuracy of mortality risk prediction, This study contribute to better healthcare practices. Accurate mortality prediction aid in identifying high-risk patients and implementing timely interventions, leading to improved patient outcomes. It support medical practitioners in making informed decisions by providing more reliable mortality risk estimates, enabling personalized treatment plans and resource allocation

## Theoretical Implications

By exploring the use of ensembling ML algorithms, the study expands the application of advanced techniques in predicting mortality risk for open-heart surgery patients, potentially leading to improved predictive models in other medical domains. This study identifies knowledge gaps and research opportunities in the field of post-operative mortality prediction. It lays the foundation for future investigations to explore additional risk factors and refine the use of ML algorithms in predicting patient outcomes after open-heart surgery.

# Conclusion

# References

Ahdal, A. A., Prashar, D., Rakhra, M., & Wadhawan, A. (2021). Machine Learning-Based Heart Patient Scanning, Visualization, and Monitoring. In 2021 International Conference on Computing Sciences (ICCS) (pp. 212-215). IEEE.

Ahmad, G. N., Fatima, H., Ullah, S., Salah Saidi, A., & Imdadullah. (2022). Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques With and Without GridSearchCV. IEEE Access, 10, 80151-80173.

Alpaydin, E. (2010). Introduction to machine learning (2nd ed.). The MIT Press.

Ashri, S. E. A., El-Gayar, M. M., & El-Daydamony, E. M. (2021). HDPF: Heart Disease Prediction Framework Based on Hybrid Classifiers and Genetic Algorithm. IEEE Access, 9, 146797-146809.

Bertsimas, D., Mingardi, L., & Stellato, B. (2021). Machine Learning for Real-Time Heart Disease Prediction. IEEE Journal of Biomedical and Health Informatics, 25(9), 3627-3637.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0 step-by-step data mining guide. CRISP-DM Consortium.

Domingos, P. (2015). The master algorithm: How the quest for the ultimate learning machine will remake our world. Basic Books.

Fitriyani, N. L., Syafrudin, M., Alfian, G., & Rhee, J. (2020). HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System. IEEE Access, 8, 133034-133050.

Garg, A., et al. (2021). "Heart disease prediction using machine learning techniques." IOP Conference Series: Materials Science and Engineering 1022(1): 012046.

Géron, A. (2017). Hands-on machine learning with Scikit-Learn and TensorFlow: Concepts, tools, and techniques to build intelligent systems. O'Reilly Media, Inc.

Gupta, A., Kumar, R., Singh Arora, H., & Raman, B. (2020). MIFH: A Machine Intelligence Framework for Heart Disease Diagnosis. IEEE Access, 8, 14659-14674.

He, Z., Zhang, H., Chen, X., et al. (2022). Hemorrhagic risk prediction in coronary artery disease patients based on photoplethysmography and machine learning. Scientific Reports, 12, 19190.

Huda, A., Castaño, A., Niyogi, A., et al. (2021). A machine learning model for identifying patients at risk for wild-type transthyretin amyloid cardiomyopathy. Nature Communications, 12, 2725.

Jindal, H., et al. (2021). "Heart disease prediction using machine learning algorithms." IOP Conference Series: Materials Science and Engineering 1022(1): 012072.

Li, J. P., Haq, A. U., Din, S. U., Khan, J., Khan, A., & Saboor, A. (2020). Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare. IEEE Access, 8, 107562-107582.

Marsland, S. (2015). Machine learning: An algorithmic perspective. CRC Press.

Niimi, N., Shiraishi, Y., Sawano, M., et al. (2022). Machine learning models for prediction of adverse events after percutaneous coronary intervention. Scientific Reports, 12, 6262.

Qian, X., et al. (2022). A Cardiovascular Disease Prediction Model Based on Routine Physical Examination Indicators Using Machine Learning Methods: A Cohort Study. Frontiers in Cardiovascular Medicine.

Ramesh, R., Kumar Lilhore, D., Simaiya, S., Kaur, A., & Hamdi, M. (2022). Predictive Analysis of Heart Diseases with Machine Learning Approaches. Malaysian Journal of Computer Science, 2022, 132-148. doi: 10.22452/mjcs.sp2022no1.10

Salah, H., & Srinivas, S. (2022). Explainable machine learning framework for predicting long-term cardiovascular disease risk among adolescents. Scientific Reports, 12, 21905.

Sarra, R. R., Dinar, A. M., Mohammed, M. A., & Abdulkareem, K. H. (2022). Enhanced Heart Disease Prediction Based on Machine Learning and χ2 Statistical Optimal Feature Selection Model. Designs, 6(2), 87.

Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge University Press.

Talin, I. A., Abid, M. H., Khan, M. M., et al. (2021). Finding the influential clinical traits that impact on the diagnosis of heart disease using statistical and ML techniques.

Wang, G., Lauri, F., & Hassani, A. H. E. (2022). Feature Selection by mRMR Method for Heart Disease Diagnosis. IEEE Access, 10, 100786-100796.

Yadav, D. P., Saini, P., & Mittal, P. (2021). Feature Optimization Based Heart Disease Prediction using Machine Learning. In 2021 5th International Conference on Information Systems and Computer Networks (ISCON) (pp. 1-6). IEEE.

Zhang, L., Wen, J., Li, Y., Chen, J., Ye, Y., Fu, Y. and Livingood, W., (2021). A review of machine learning in building load prediction. *Applied Energy*, *285*, p.116452.