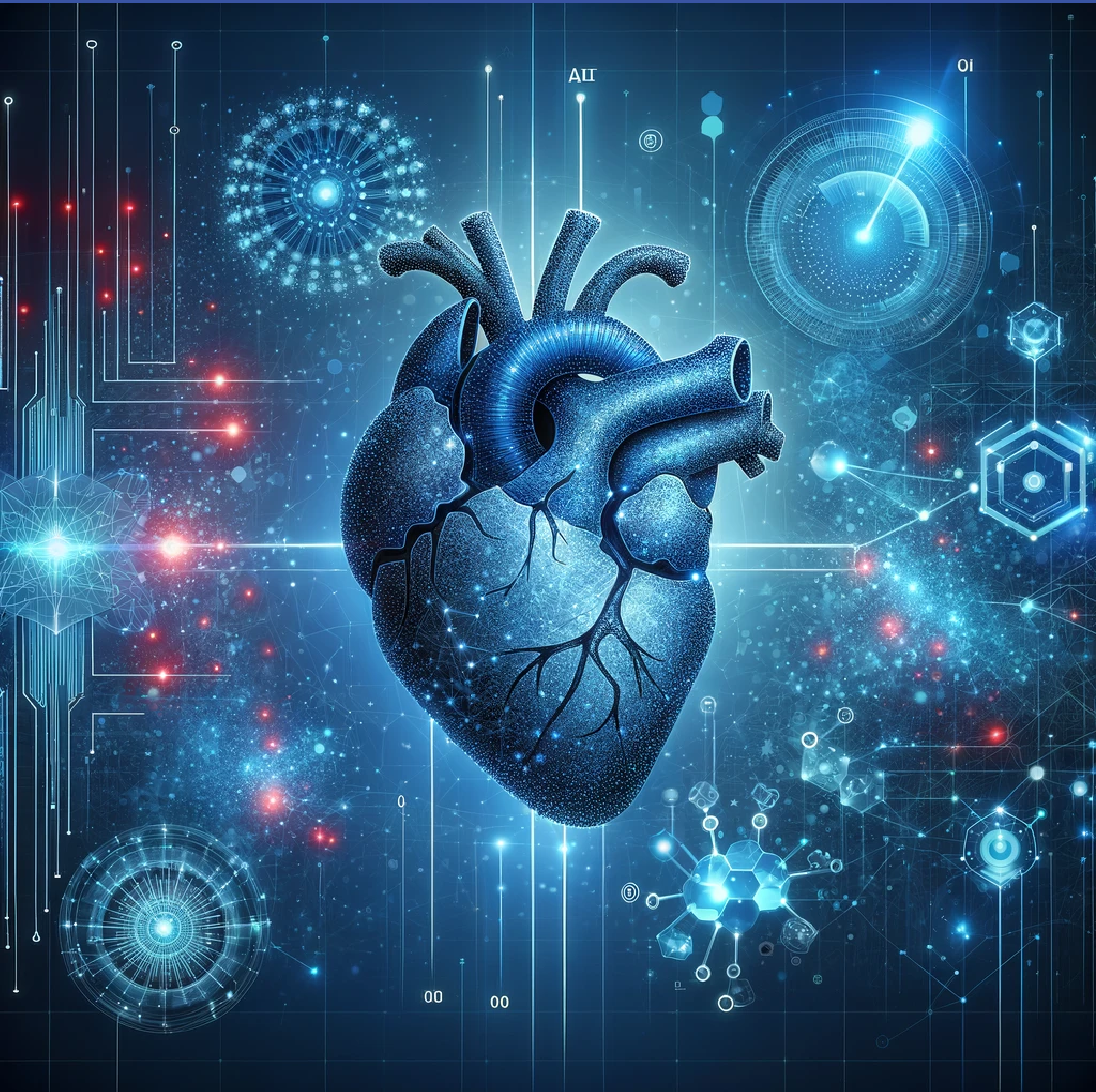
DTU Compute
Department of Mathematics and Computer Science

Bachelor of Artificial Intelligence and Data
Organ failure following cardiac surgery

02466 - Project Work



Bachelor of Artificial Intelligence and Data
Organ failure following cardiac surgery

02466 - Project Work
Date, 2024

By
Ditte Bjerrum Gilsfeldt *(s210666)*
Lucia Han Lu *(s224215)*
Muneer Kayali *(s214642)*
Viktor Skovlykke Sølvsten *(s225784)*

Cover photo: Brendon Hobson, January 23, 2024
Unpublished by: DALL·E 2024-03-24 12.43.09, (9), Create an abstract illustration that embodies the concept of artificial intelligence surpassing traditional logistic regression methods in predicting .webp
https://chat.openai.com/c/bd763d60-999d-4eb1-a1b7-85aa7e802882

Approval

Ditte Bjerrum Gilsfeldt - *(s210666)*

....

Signature

....

Date

Lucia Han Lu - *(s224215)*

....

Signature

....

Date

Muneer Kayali - *(s214642)*

....

Signature

....

Date

Viktor Skovlykke Sølvsten - *(s225784)*

....

Signature

....

Date

Abstract

Acknowledgements

Ditte Bjerrum Gilsfeldt ***(s210666@dtu.dk)***, BSc. Artificial Intelligence and Data, DTU
Lucia Han Lu ***(s224215@dtu.dk)***, BSc. Artificial Intelligence and Data, DTU
Muneer Kayali ***(s214642@dtu.dk)***, BSc. Artificial Intelligence and Data, DTU
Viktor Skovlykke Sølvsten ***(s225784@dtu.dk)***, BSc. Artificial Intelligence and Data, DTU

[Morten Mørup], [Professor], [Guidance counselor]

[Lars Grønlykke], [MD, PhD], [Guidance counselor]

[Chaoqun Zheng], [MD,PhD], [Guidance counselor]

[Theis Skovsgaard Itenov], [MD, PhD], [Guidance counselor]

Contents

1 Preface

This project, Organ failure following cardiac surgery – the use of AI for prediction modeling, is written by a group of four bachelor students from the Technical University of Denmark (DTU) with the field of study being Artificial intelligence, interested in machine learning. Moreover, this project is built upon the interest regarding how such technology is used in the vast sector of healthcare, which is found to be fascinating. This journey is, therefore, about exploring how the potential of artificial intelligence for prediction modeling of organ failure following cardiac surgery is used. This will also act as a personal commitment, experience and meaningful growth.

The importance of this project lies within showing how the technology of machine- and deep learning can be useful for predicting morbidity and mortality, which includes organ failure following cardiac surgery. In addition to this, the hospitals are highly interested in optimizing the number of patients surviving the surgery, because in worst cases, patients will end up in intensive care unit (ICU), or even end up deceased. Hence, the hospitals are interested in assessing if there exists a correlation between certain medical factors, which is to determine if a patient is suited for surgery or not in early manners. Therefore, this project hopes to build models that can guide the hospital when making clinical decisions. Such models can provide information about the risk of the surgeries for different patients, and, further, to optimize the ICU, so hospitals can attain correct resources post-surgery. In summary, this project aims to provide the hospitals with tools for predicting morbidity and mortality, which can help optimize the survival rate, usage of certain resources and to minimize complications post-surgery (3).

Look forward to the future, these models will help the hospitals to create more precise predictions and detect the complications earlier.

Bibliography

[1]

[2] Casale Pierluigi, Pujol Oriol, R. P. *BOOK CHAPTER · JOURNAL ARTICLE - Embedding random projections in regularized gradient boosting machines*. Springer Berlin Heidelberg, 2011, p. 203.

[3] Grønlykke, L., and Theis Skovgaard Itenov (Rigshospitalet), M. M. Organ failure following cardiac surgery – the use of ai for prediction modelling. ?

[4] Kanade, V. What is logistic regression? equation, assumptions, types, and best practices, 2022.

[5] Kobayashi, Y., Peng, Y.-C., Yu, E., Bush, B., Jung, Y.-H., Murphy, Z., Goeddel, L., Whitman, G., Venkataraman, A., and Brown, C. H. Prediction of lactate concentrations after cardiac surgery using machine learning and deep learning approaches.

[6] Lek, S., and Park, Y. S. Multilayer perceptron. 2455–2462.

[7] Manaswi, N. K. *Deep Learning with Applications Using Python Chatbots and Face, Object, and Speech Recognition With TensorFlow and Keras*. Apress Berkeley, CA, 2018, pp. 47 – 49, 115 – 126.

[8] Nashefa, S. A., Roquesb, F., Sharplesc, L. D., Nilssond, J., Smitha, C., Goldstonee, A. R., and Lockowandtf, U. Euroscore II, 6 January 2012.

[9] OpenAI. Dall·e 2024-03-24 12.43.09 - create an abstract illustration that embodies the concept of artificial intelligence surpassing traditional logistic regression methods in predicting. 24 Marts 2024.

[10] Paluszek, Michael, Thomas, Stephanie, Ham, and Eric. *Practical Matlab Deep Learning*. Apress, 2022, ch. What Is Deep Learning?, pp. 1–24.

[11] Panesar, and Arjun. *Machine Learning and Ai for Healthcare*. Apress, 2021, pp. 62–83.

[12] Skansi, S. *Introduction to Deep Learning - From Logical Calculus to Artificial Intelligence*. Springer International Publishing AG, 2018, pp. 136–137.

[13] Templ, M. *Statistical Disclosure Control for Microdata*. Springer International Publishing, 2017, p. 157.

[14] Yang, T., Yang, H., Li, Y., Lui, X., Ding, Y.-J., Li, R., Mao, A.-Q., Huang, Y., Li, X.-L., Zhang, Y., and Yu, F.-X. Postoperative delirium prediction after cardiac surgery using machine learning models, February 2024.

[15] Zheng, C., Itenov, T., , and Grønlykke, L. *Cardiac surgery prediction modelling*. 2024, pp. 3, 4.

2 Introduction and Motivation

When patients undergo cardiac surgery, a serious number of risks are known to follow. This includes declining health, complications during surgery, sicknesses, and in worst cases, mortality. Hence, numerous research fields are focused on measuring such risks with an aim to reduce and, ultimately, prevent them. The most widely used tools for estimating complication rates and risks are EuroSCORE II and STS risk score. However, these risk scores come with shortcomings, as they lack precision in estimating surgery risks for the individual patient, and the models are designed to predict mortality but not morbidity. Hence, it is relevant to research the possibilities of using Artificial Intelligence (AI) to create stronger prediction models, compensating for the weaknesses observed in the current models. In such researches, EuroSCORE II will act as a baseline model for comparing different learning models that will be developed (3). Ultimately, these models, in opposition to existing ones, will be trained to detect early complications in patients aiming to help prevent mortality and, especially, morbidity. It is known that out of 1800 yearly patients who undergo cardiac surgery at Rigshospitalet in Copenhagen, a crucial amount of these individuals will experience complications post surgery, keeping them hospitalized longer, and sometimes their health becomes fatal. Therefore, the investigation done in this project is essential, as this newfound knowledge can serve as guidance for doctors and patients deciding on the surgery (3).

To summarize, this project will investigate the possibilities of using AI to create stronger prediction models, and it will contain Machine Learning (ML) and Deep Learning (DL), both representational for AI. By using these AI techniques, specific features can be found to provide patterns for analyzing specific situations of patients. Moreover, the ML models will provide a dynamic predictive tool, which is useful for making decisions regarding, if a patient should undergo surgery (14). In addition to this, it is known that there already exists an ML model for short term mortality, namely the mentioned EuroSCORE II model. This project will be applying a more complex model to yield a stronger prediction ability regarding both short- and long-term mortality, with the ultimate goal of achieving a higher accuracy. This will be determined through a Random Forrest model (RF). Furthermore, a Multilayer Perceptron model (MLP) will be useful for predicting the mortality of cardiac surgery. This model will also be useful for recognizing patterns and possible correlations.

It is hypothesized that a deep learning model has the potential to outperform EuroSCORE II for accurately predicting the risks of organ failure (morbidity) and mortality after cardiac surgery. It is known that deep learning models are specialized in complex relations, and, hence, the goal is to predict that such models will be learning subtle correlations and non-linear dependencies within the complex dataset that would, otherwise, be overseen by a simple logistic regression model as the one that was used to create EuroSCORE II. Lastly, the data set used in this project is based on personal numbers, hence, it is protected under circumstance of the GDPR legislation. Consequently, this project will be utilizing all information and conclusions based on the original data set, but also a synthetic data set.

This project will be elaborated through a state of the art section, describing what has previously been found concerning this topic, a section for data description, theory and method and, finally, a results, analysis and discussion section rounded up by a final conclusion.

3 State of the art

This section describes existing academic theory concerning previously used methods for creating the EuroSCORE II model, which predicts risks of mortality of patients undergoing cardiac surgery, along with other improved machine learning models based on the same field. These existing theories motivate the methods that this project uses for research.

Firstly, theory of the EuroSCORE II model, based on the belonging EuroSCORE II article (8), will be described. The goals of this article were to update the European System for Cardiac Operative Risk Evaluation (EuroSCORE) risk model that originated from the year 1995. This model is mentioned as the first EuroSCORE model. However, in this article, the newer EuroSCORE II model, from 2012, was developed and trained based on data collected from 22381 patients that had undergone major cardiac surgery. The data was collected over a 12 week period, from May to July, in 2010. It was also mentioned that along with the risk factors already considered in the existing EuroSCORE model, new risk factors had been identified through the article's research, providing a more comprehensive understanding of the risks involved in cardiac surgery when creating EuroSCORE II. Moreover, for training this model, data was split into a subset for training a series of single variable logistic regression models and a subset for testing the models. The article concluded that EuroSCORE II was superior, as it was better calibrated while preserving powerful discrimination. In addition to this, the article argued that this superiority was based on the fact that the models were trained on current data set, thus, being better at reflecting current cardiac surgical practice. Finally, it was also deduced, from the models, that cardiac surgical mortality has been significantly reduced in the last 15 years despite including older and sicker patients.

The methods used in creating EuroSCORE II served as heavy inspiration in previous studies that attempted to fit a logistic regression model onto the exact data set that this project explores. However, apart from studying and applying logistic regression models as EuroSCORE II, it is mostly of interest to create stronger machine learning models for determining complications post surgery and predicting mortality. This implies developing models that covers what EuroSCORE II is unable to, which is the main goal of this project. Previous studies have shown to utilize deep learning and machine learning methods. The article *"Prediction of lactate concentrations after cardiac surgery using machine learning and deep learning approaches"* (5) is an instance of this, since it highlights how researches have used machine learning models as random forest, artificial neural network, and a multivariate linear regression model with an focus on predicting lactate concentration in patients post cardiac surgery. Higher levels of lactate concentration were described as a complication post surgery. One of the key factors in this article was that it argued that numerous prediction models in cardiac surgery settings only used static patient characteristics and a limited number of intraoperative variables. However, it was known that during cardiac surgery, particularly during cardiopulmonary bypass, minute level data was available on key parameters such as flow, hemoglobin concentration and mean arterial preassure. Because conventional prediction models often ignored such dynamic data, deep learning approaches were better suited to incorporate time varying and non linear data that had more complex interactions. This knowledge of deep learning being able to take advantage of minute level data is serving as important influence on methods that will be used throughout this project when developing new machine learning models. Besides, in contrast to the article, this project will not be limited to predicting only one complication, but numerous, as it is focused on developing models for predicting general post-operative complications, such as organ failure and, essentially, morbidity.

4 Data description

The data used in this project, is collected from 8000 patients over five years, who have undergone cardiac surgery at Rigshospitalet. The Cardiac patient data journey has a Preoperative section such as; ”sex, age, medical- and surgical history etc.”, Perioperative as; ”blood pressure, heart rate, pressure, etc.”, ICU as; ”blood pressure, heart rate, dialysis, etc.”, and Ward as; ”blood pressure, heart rate, medication” from appendix A.1.

There will be a restriction when working with the data, due to the GDPR-rules, thus, the data can only be used on one specific device when training and creating the machine learning models. This also means that all personal information has been anonymised, removing information like CPR numbers.

Each measurement and test result is linked to a specific identification number on the form ”ID 1” in all data files, assuring the tracking of correlation between the samples. None of the observations is specifically labelled as preoperative, perioperative or postoperative in practice. Measurements like blood pressure are often sampled once under all three categories. Instead, each observation has, along with the ID, a date and time that describes when each observation was noted. This makes it possible to keep track of when in the surgery process of a patient the observations are made. for a general understanding of the dataset, the following categories have been made:

4.1 Profile

The profile describes information about the patient, that is already known before the process of operation begins. The variables documented are:

Profile variables	
Biological gender	Male or Female
Age	Years
Diagnosis	Name and SKS
Alcohol	Drinks per week
Smoking	Frequency and previous history

Table 4.1: Profile variables

4.2 Measurements

This section of the data describes all the information that is collected during the patient’s hospital duration. These data contain the most information about why and what future complications the patients will experience. In other words, these data will most likely make up the majority of the predicting variables.

Measurement variables	
Blood pressure	Millimetres Mercury(mmHg)
Mean Arterial Pressure	mmHg
Invasive Arterial blood pressure	mmHg
Central venous pressure	mmHg
Pulmonary Artery Pressure	mmHg
Oxygen saturation in the blood	Percentage
respiratory rate	Breaths per minute
Blood and body temperature	Celsius
Pulse	Beats per minute
Weight	in Kilograms, measured several times over the process
Urine	Millilitres
Laboratory test results	Very broad, but most important is liver health and infections
Medicine administered	Both name and amount
EKG results	Several measurements
echocardiography	Several measurements
Coronary angiography	Examination of Coronary arteries
Spirometer test	Ventilation and general health of the lungs
Other operations performed	A week after the cardiac surgery
Total incretion	Food, liquid and IV drop ETC.
Total excretion	Defecation, urine, vomit, bloodloss ETC.
Ventilator data	Duration of the assisted breathing and settings
Duration of anaesthesia	Describes the duration of cardiac surgery
anaesthesia data under the operation	Several measurements
Blood loss during operation	millilitres and cause of bloodloss
CV bypass and closing of aorta	Duration of the manoeuvre

Table 4.2: Measurement variables

The most important variables to understand from the dataset would be the CV bypass and closing of the aorta, which is how the blood is kept circulating through the body during the operation. The duration of this can often describe how successful the operation was and how much the patient will bleed during and after the operation, which is a key factor in recovery time.

4.3 Admission process and follow up

The Admission process describes how long each patient spends in the different locations. This is documented by specifying the location, for example ”Thorax intensiv” and also the start of admission and end of admission. Sometimes a patient is transferred from one ICU to another for logistical reasons but as the care in the different ICUs is the same this can be ignored. The duration spent in the ICU will be one of the main target variables

The follow-up information with the patients is also documented. This includes variables such as if death has occurred within a year and if so when. This is a very important variable for describing the overall success of the operation.

5 Research questions

The main focus is to assess the risk of either mortal or non-mortal complications related to cardiac surgery, especially to investigate the risk of mortality after 30 to 90 days post-surgery. Another area of interest is the prediction of time spent in the ICU and other complications following the cardiac operation, which lies under morbidity.

The following research questions have been formulated:

- To what extent is a Machine Learning (ML) model able to outperform the EuroSCORE II, in predicting the mortality of a cardiac surgery patient, and is it more practical?
  - The goal of this question is to create a model that can give a higher accuracy than the EuroSCORE II, or conclude that the current EuroSCORE model is better suited for predicting mortality following cardiac surgery.
  - This will be operationalized by first choosing and designing a suitable ML model, that can predict the probability of mortality, like the EuroSCORE II. It is important to create a common ground for comparison, for a valid result.
  - To outperform the EuroSCORE II, an RF ML model, could be a powerful method to use, in fact, RF is useful to handle data, that is non-linear, which can help to understand the influence of mortality. Further, could a Multilayer Perceptron (MLP), also be useful, for the reason that it can capture patterns that can have an influence on mortality better than a logistic regression.
- If there exist correlations between different groups and the outcome of the surgery, how can they be detected using AI and what influence would it have on predicting morbidity?
  - The goal of this question is to determine if there is a correlation between the predicting variables and the outcome of the surgery. A potential conclusion would be that there is a clear correlation between a rise in blood pressure after the surgery and an extended stay at the ICU.
  - This will be investigated by using either a regression or classification model, depending on which model makes the most sense. Certain groups will be chosen with either logical reasoning or, methods like a MARS model.
  - To find a correlation between the different groups and the outcome, a RF model can might be used to provide insight into the correlation, moreover can a Logistic Regression with Interaction Terms to analyze the impact of the variables, to find out if the patient should go on surgery or not.

- How well can ML models predict post-operative complications, like organ failure, to prevent patients from staying in the ICU?
  - The goal of this question is to create an AI model that can determine the expected length, a patient will spend in the ICU and what causes this stay in the ICU.
  - This question will be operationalized, by firstly creating some relevant metrics to predict. Afterward, a Deep Learning Network will be fitted to the data to find linear or nonlinear correlations, between measurements and the chosen post-operative metrics.
  - To predict the post-operative complications, a ML model, Gradient Boosting Machines (GBM), can be used to handle an unbalanced dataset which would be typical for dataset like this, and the model is good to handle missing data as well. While adjusting the loss function the GBM model can be tuned in to focus on the high-risk complications. Further a RNNs or LSTM Networks can be used to capture patterns that might predict complications.

6 Theory

This section will be a draft of potential theory.

6.1 EuroSCORE II:

EuroSCORE II is a risk assessment tool used in cardiac surgery to predict the probability of mortality following cardiac procedures. This model is an update to the original EuroSCORE model. EuroSCORE II is trained on more variables than its predecessor, including a set of preoperative, intraoperative, and postoperative variables.

Table 2: Data set
Patient-related factors
Age and sex
Height and weight
Pulmonary disease
Diabetes status
Extracardiac arteriopathy
Neurological or musculoskeletal dysfunction
On dialysis
Last serum creatinine
Brain-natriuretic peptide
Serum albumin
Cardiac-related factors
Symptomatic status
NYHA
CCS
LV function
Recency and size of last myocardial infarct
Systolic PA pressure
Active endocarditis
Previous cardiac surgery
Operation-related factors
Urgency
Elective
Urgent
Emergency
Salvage
Type of procedure(s) performed in detail
Times of
Bypass
Cross-clamp
Deep hypothermic arrest
Selective cerebral perfusion

Figure 6.1: Variables used in training the EuroSCORE II model

These variables can be seen in figure 6.1. The algorithm of EuroSCORE II operates on a logistic regression model, where a linear combination of the input variables shown in 6.1 is transformed into a probability of mortality using a sigmoid activation function.

EuroSCORE II is widely used around the world as the primary risk assessment tool for predicting mortality following cardiac surgery. While EuroSCORE II has been a valuable tool in risk assessment for cardiac surgery, there are limitations to its predictive accuracy, particularly in capturing complex nonlinear relationships and subtle interactions between variables (8).

6.2 Machine Learning:

Machine learning (ML) is a subset of AI where computer models are trained to learn from their actions and environment over time with the intention of improving their performances (11). Such models are able to find certain patterns in given data, allowing them to perform better, the more they learn.

6.2.1 Logistic Regression (LR):

The logistic regression model is used for binary classification tasks by predicting the probability of an outcome, event, or observation. This model analyzes the relationship between one or more independent variables to classify data into discrete classes. Furthermore, it is mostly used in predictive modeling, where the model estimates a mathematical probability of whether an instance belongs to a specific category or not (4).

6.2.2 Gradient Boosting Machines (GBM):

Gradient Boosting Machines (GBM) is a regression and classification problem used to find a function $F^{*}(x)$, so the expected value of the loss function will be minimized when the function maps x to y over a joint distribution for all (y, x)-values. Normally the function $F(x)$ will be a parameterized class of the function $F(x; P)$ where $P = \{\beta_m, a_m\}$ is a parameter set (2).

6.3 Deep Learning:

Deep learning (DL), a subset of machine learning, refers to neural networks with more than one layer of neurons. Neural networks behave similarly to the neurons in the human brain, therefore, the name ”deep learning” is taken to imply that a learning system is a deep thinker. In addition to this, since deep learning networks have multiple layers before the final output, they are able to solve more complex problems, especially regarding predictions and classifications (10).

6.3.1 Multilayer Perceptron (MLP):

The Multilayer Perceptron (MLP) model is generally using an ANN class for non-linear modeling (6). MLP is a neural network with an input layer, at least one hidden layer and an output layer. The model is a non-linear activation function like sigmoid or ReLU. The method is a supervised learning technique for training, where the loss function, such as entropy, will be minimized. This is called *backpropagation*. This technique uses an optimizer for controlling the parameters (weight and bias). Since it is calling *backpropagation* meaning that the errors will go from the output to the hidden layer then to the input layer, so the learning rate and optimizer will be regulated by adjusting the parameters (7).

6.3.2 Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) Networks

Recurrent Neural Networks (RNN) are extremely powerful algorithms used for classifying, clustering, and making predictions about data, particularly time series (1). In RNNs there are three settings, classic setting, sequential setting and predict-next setting. The classic setting is the Naïve Bayes classifier which is calculated by $P(\text{target}|\text{features})$. The RNN model has multiple labels and can learn through a lot of labeled sequences. Following that, it can predict labels of all finished sequences. The sequential setting is that the RNN can learn from uneven sequences, where some parts are labeled and then create new labels. The predict-next setting does not need labels. This setting takes every input sequence of the RNN and downgrades it to subsequence (12).

Long Short-Term Memory Networks (LSTM) are networks based on a modified version of the RNN architecture. LSTMs are able to address the problem of training over long sequences, as these networks have a memory cell that can maintain information for long periods of time. On the contrary, RNNs have difficulty learning through longer time periods due to the gradient of the loss function decaying exponentially through time, also called the vanishing gradient problem. However, LSTMs uses a set of gates to control when information enters the memory to solve the vanishing gradient problem (7).

6.4 Synthetic Data:

The synthetic data require that the transformation of the data will be created as a new anonymous micro data. Furthermore, the synthetic data must be realistic and be proportionate with the actual population. The reason for using synthetic data is due to the disclosure risks being much lower than the method, where the data is anonymized. The intention for the synthetic data is not to replace the traditional data set, but rather to create a replacement for the dataset, so it can be shared to the public (13).

7 Methods

This section will be a draft of potential methods.

7.1 Clean Data

The first thing to do before applying machine learning and AI models to the data is to clean the data. This will be done by handling outliers, missing data and unwanted observations. The data is very messy, structured in multiple different tables that hold duplicate observations, unconventional column namings and inconsistent metrics used. Duplicates or observations deemed either unnecessary, irrelevant or lacking sense will be removed to reduce noise and improve the overall quality of the data. Furthermore, outliers will be identified, and either be removed or transformed to minimize the impact these will have on the model. Finally, the different tables that the data comes in will be merged into one big dataset. This will require making decisions about which columns to merge together or to even include in our final dataset. This whole stage of data cleaning will require a deep dive into the individual observations of each table, such that one can analytically assess the quality of these and decide what to do accordingly.

7.2 Feature selection

At this stage, the relevant features for the model will be chosen. This is arguably the most important part of any machine learning and AI model and could be the factor deciding whether a model accurately predicts the wanted output or not. To do this, the columns that are deemed irrelevant or could not possibly hold any information that would help the model predict wanted problems, for instance mortality and morbidity of the patients, will be removed from the dataset. To achieve this, the doctors from Rigshospitalet will be consulted, who will help assess the relevancy of different features. It is expected that many columns included in the raw data will not be present in the final dataset, resulting in considerably smaller but higher quality feature space. Once this is done, the data will be ready for preprocessing.

7.3 Preprocessing

In this stage, the data will be transformed such that is a better fit for machine learning analysis. This will done by changing some of the metrics or formats of some of the features such as ”time” which is observed in strings containing the date and hour to something more numerical. The goal here is to try to transform the whole dataset such that we avoid non-numerical values as much as possible where it is relevant. This can be done by one-hot encoding and many different methods which are yet to be discussed. Then, all the numerical variables will be normalized, such that the distribution has a mean close to 0. This generally helps machine learning models learn more efficiently and converge faster. Considering the messy format of the data at hand, this stage will require many sophisticated data science methods where it will be necessary to analyze the different features deeply to assess which methods should be applied.

7.4 Synthetic Data:

The data at hand is very sensitive as it includes personal information. For this reason, it will be necessary to use synthetic data to be able to easily access it from different Python devices. This will make it possible to run multiple different models on different devices at the same time, which will help in the assessment of deciding which models work best on the type of data at hand. Once the data is all clean and ready for training, a machine learning model (GAN) will be applied to learn the distributions of all the different variables. This machine learning model will then be used to generate new data points that pull from the same distribution as accurately as possible. These new data points will be used to create new datasets that can be trained on. This can be done by using different Python libraries such as data-synthetic or sci-kit learn. The exact method that will be used to create synthetic data is yet to be evaluated.

7.5 Logistic Regression and Random Forrest:

One model, within the machine learning branch, that will be tested will be logistic regression. The model should output a percentage of mortality following cardiac surgery. This will be the baseline model as EuroSCORE II is trained with the same algorithm. Different hyperparameters will be tested, and the best will be chosen. Doing this will give a good idea of how the dataset at hand competes with the data that EuroSCORE II is trained on. This will also give insight towards whether the dataset should be preproccessed differently or whether or not different sets of features should be tested. Once this is done, the model will be saved and used for comparison with other models in the following stages. Another model that will be implemented is a random forest models. The hypothesis is that the data at hand contains complex non-linear relations. This type of model could outperform the classic logisitic regression for this reason. Other models within machine learning, for instance GBM mentioned in the theory section, will not be further explained, since this method has not been explored yet.

7.6 Deep Learning and training:

The main model that will be implemented and refined will be a deep neural network. This neural network should contain many hidden layers, as the feature space of the data will likely be relatively high dimensional. The depth of the neural network will help capture relatively complex relations and patterns. The output of the neural network will be multidimensional, including a percentage of mortality/morbidity and/or risk of different types of organ failures among other things. The exact nature of the output is yet to be determined. Furthermore, the depth of the neural network is still unknown, however the goal is to make the neural network be as deep as possible without largely compromising training time, as there is no access to super computers. Once this model is refined and the exact nature of it is determined, it will be ready for training and testing. The model will be trained using one of the methods available in the python packages that will be used, which will likely be the Adam optimizer. Once the model is trained, it will be ready for comparison. It is worth to note that thorough choices of specific type of neural network and whether or not the use of RNN's or LSTM's is relevant for this task is not yet clear, as data has not yet been worked with.

7.7 Statistical evaluation and model comparison

Finally, the DNN model will be run through different statistical tests to determine it's accuracy. The exact tests that will be used will depend on the nature of the model's output. Then the model will be compared to the logistic regression model that was trained on the same data and the EuroSCORE II model. The method used to compare these models will depend on many different factors, including the DNN's output nature and the nature of the data fed into the models among other things. It will be difficult to compare the neural network model with the EuroSCORE II model as the data fed into these model have different feature spaces. This will likely pose a problem which will later be discussed more deeply once a foundation is built, but the way of comparison could end up being slightly subjective given the nature of the task at hand.

A Appendix

A.1 Pictures from slide

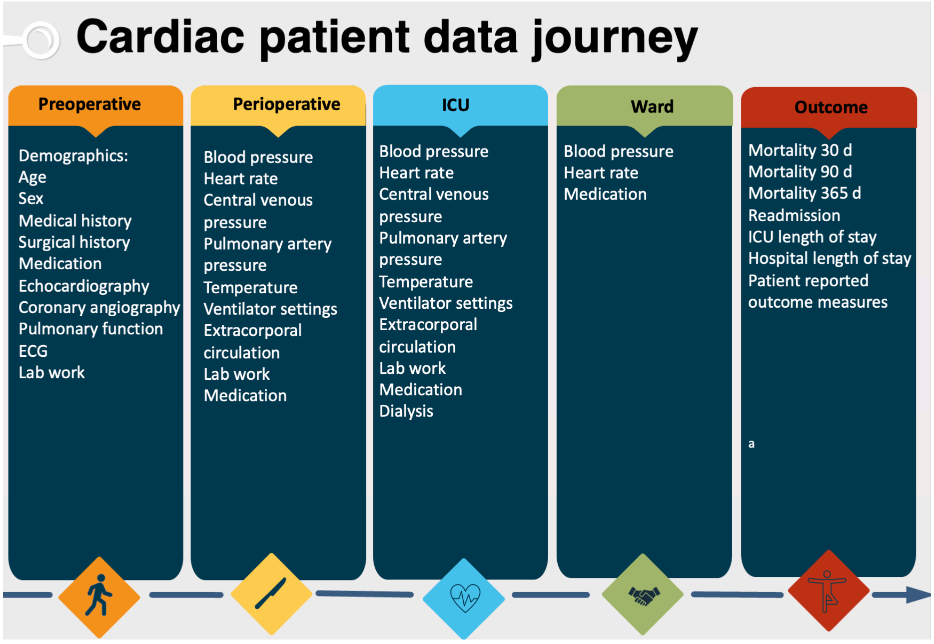


Figure A.1: Cardiac patient data journey (15)

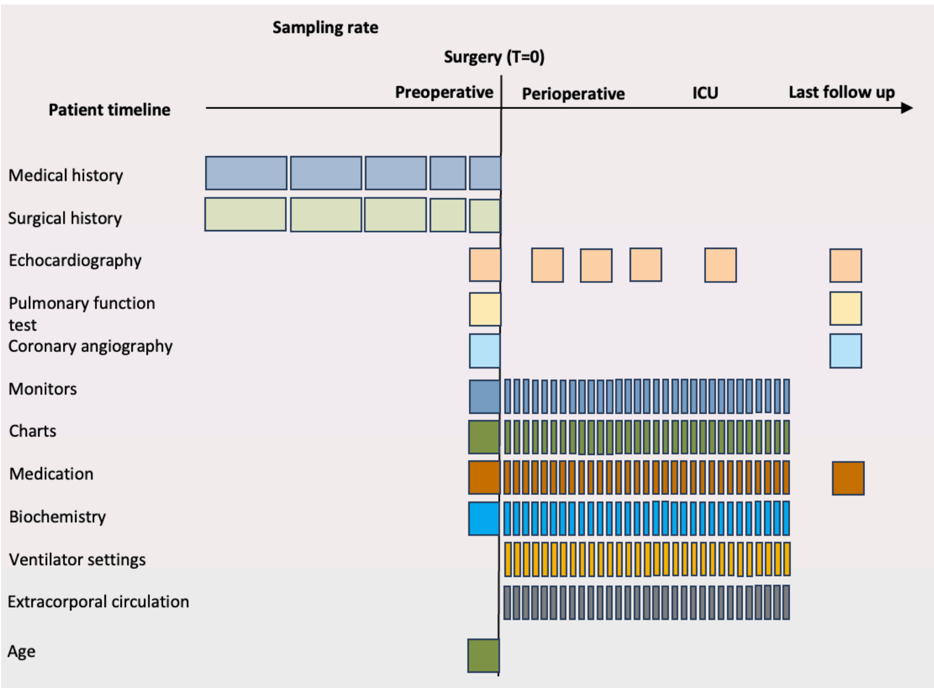


Figure A.2: Sampling rate (15)

Technical
University of
Denmark

DTU Compute Bygning 324
2800 Kgs. Lyngby
Tlf. 4525 1700

https://www.dtu.dk/uddannelse/bachelor/uddannelsesretninger/kunstig-intelligens-og-data