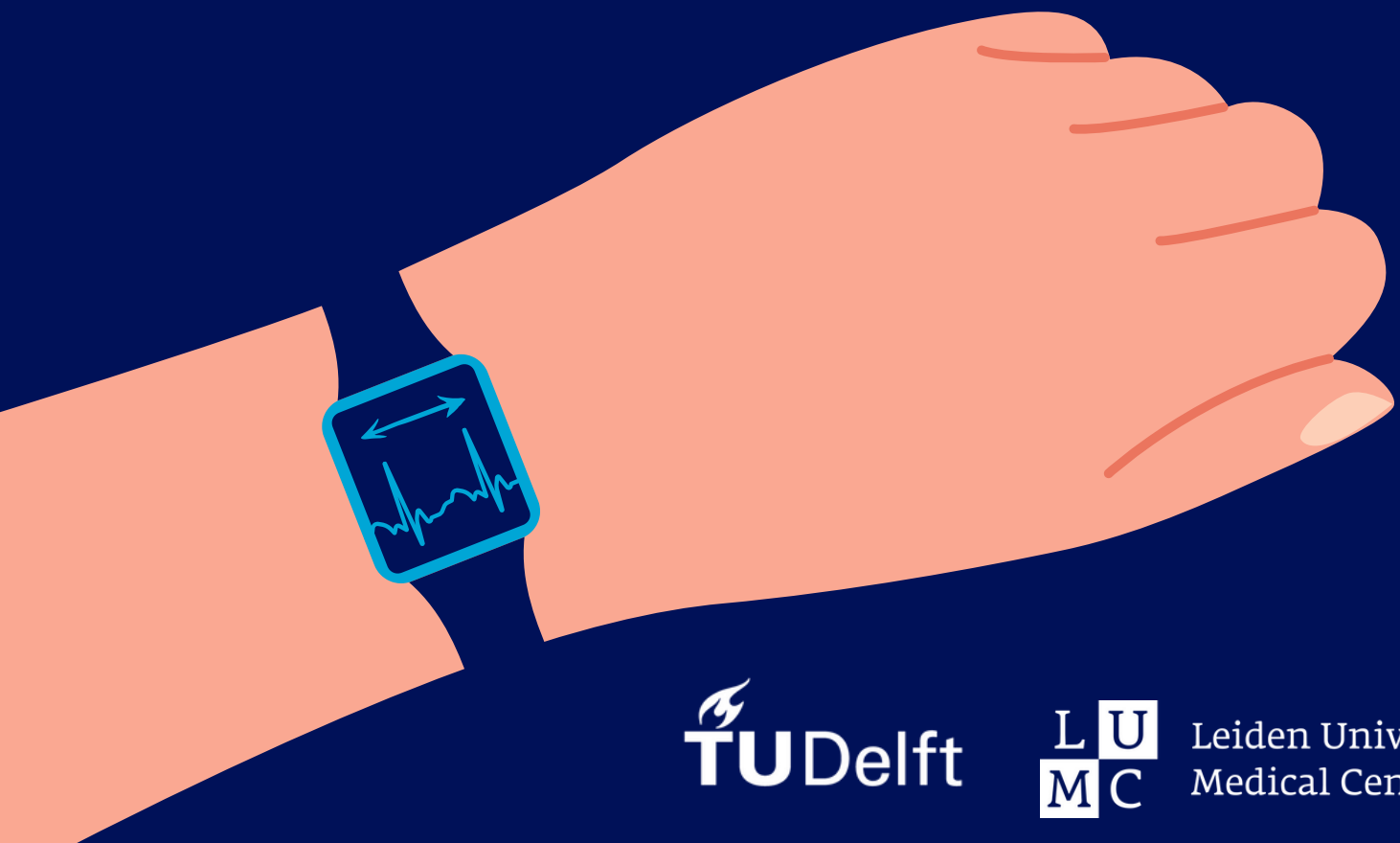
# Improving Intensive Care Unit outcome using Heart Rate Variability-based Machine Learning analysis and eHealth monitoring

*Puck Noorlag*

August 2024

Master thesis Technical Medicine



TU Delft

LUMC Leiden University Medical Center

This page was intentionally left blank.

# Improving Intensive Care Unit outcome using Heart Rate Variability-based Machine Learning analysis and eHealth monitoring

Puck Noorlag

TM30004 MSc Thesis

Student number: 4560833

August 14, 2024

Thesis in partial fulfilment of the requirements for the joint degree of Master of Science in

*Technical Medicine*

Leiden University | Delft University of Technology | Erasmus University Rotterdam

Dept. of Biomechanical Engineering, TU Delft
Department of Intensive Care, LUMC

*December 4, 2023 – August 28, 2024*

Supervisor(s):
Prof. dr. M.S. Arbous
Prof. dr. ir. M.J.T. Reinders

Thesis committee members:
Medical supervisor and chair: Prof. dr. M.S. Arbous, LUMC
Technical supervisor: Prof. dr. ir. M.J.T. Reinders, TU Delft
Independent member: Dr. D.M.J. Tax, TU Delft



Universiteit Leiden

TU Delft Delft University of Technology

Erasmus Universiteit Rotterdam

# Preface

With this thesis, I conclude my time as a student. It all began in 2016 when I started the Clinical Technology bachelor program. During my time in Delft, Leiden and Rotterdam, I have not only had the pleasure of following the Clinical Technology and Technical Medicine study program, but I continued the med tech theme in other parts of my student life. I chose a minor that focused on med tech-based entrepreneurship, and after obtaining my bachelor's degree, I spent a year as part of the TU Delft D:Dreamteam Project MARCH.

I was a part of the third year in which the study program was offered in Delft. And that did not go unnoticed; when I told someone what I was studying, they often did not understand because it was still a very unknown and unexplored field. And to be honest, I did not really understand it that well either. I had initially chosen it because I found the human body interesting and was good at math. The standard answer given to us during the program was that we were the bridge between the doctors and the engineers. I repeated this for years to anyone who would listen. And now, after six and a half years of studying, I have not only come to believe it, but I based my MSc thesis on this principle by conducting two studies. One study focuses on applying my technical skills to solve a medical problem, while the other is a clinical study where my technical and analytical skills proved particularly useful. These two studies have been translated into two chapters that can be read independently of each other.

Before I wish you an enjoyable read, I would like to thank a few people. First and foremost, my supervisors. Marcel, thank you very much for your sharp insights and contributions to Chapter 1. Thanks to you, we decided to explore the correlations, after which the structure of the story became very clear. Thank you for allowing me to visit you to discuss the key principles of Machine Learning and for your patience in explaining them. Sesmu, I would like to express my deep gratitude for all the time and effort you have put into me and my thesis over the past nine months. You have provided immense support in both parts and were always available for brainstorming or advice. I always felt energized after our meetings. I would also like to thank Tina and all the TM2 students who contributed to the realization of Chapter 2. I always looked forward to our Wednesday afternoon meetings and enjoyed taking on a sort of supervisory role.

Above all, I want to acknowledge the support I received from my friends and family. Thank you all for your support during the writing of this thesis. In particular, I want to thank my roommates, without whom my evenings and weekends would have been much less enjoyable. Appie, thank you for always being there for me. The same goes for my boyfriend, Pim —thank you for your unconditional support. Lastly, and most importantly, a huge thank you to my parents. Mom and dad, without you, I would never have been able to complete my studies in this way. I will always be grateful for this.

That leaves me with nothing more to say than to wish you an enjoyable read of my MSc thesis!

*Puck Noorlag*
*Rotterdam*
*August, 2024*

# Summary

The intensive care unit (ICU) is a hospital department where critically ill patients requiring organ support or intensive monitoring are admitted. Nowadays, the care provided in an intensive care unit has advanced so that more patients are being discharged alive. Advances in ICU care have increased patient survival rates, yet ICU admission is still associated with high morbidity and mortality both during and after the stay.

Identifying patients at high risk of complications during ICU admission is crucial. Even though there has been an increasing focus on predictive models, making accurate predictions with the data currently available remains challenging. A non-conventional parameter that appears to have promising value is heart rate variability (HRV), which reflects the fluctuations in time intervals between consecutive heartbeats and can be derived from the electrocardiogram. We investigated whether heart rate variability was able to predict ICU mortality and ICU length of stay. We employed a machine learning approach, assessing three models for each outcome. We used nested cross validation to estimate the performance and optimize and select the final models. Data from 468 adult patients admitted to the ICU for 48 hours or longer were analyzed and nine HRV measures were calculated. Two HRV measures, the power in the high frequency band and the standard deviation of 5-minute average RR intervals, showed a significant difference between the ICU survivors (n=398) and ICU non-survivors (n=71). While individual HRV measures had limited predictive power for ICU mortality, combining HRV with clinical features improved performance. The best performing model was an eXtreme Boosting Gradient classifier that used clinical features in combination with three HRV measures (power in the high frequency band, power in the low frequency band and the ratio between the power in the low frequency and high frequency band) achieving an AUC of 0.76. The models predicting ICU length of stay performed poorly, with the best model achieving a mean absolute error of 5.07 days. These findings suggest that HRV, when combined with clinical features, has potential in predicting ICU mortality, though further research is necessary before clinical implementation.

Monitoring ICU survivors after discharge is equally important, as half of these survivors suffer from Post Intensive Care Syndrome, which negatively impacts their quality of life and increases their healthcare needs. A team of researchers from the ICU conducted a pilot study establishing the feasibility of monitoring fifteen ICU survivors using eHealth, but a larger clinical trial was needed to assess feasibility on a larger scale. Therefore, we aimed to prepare for the ICU Recover Box 2.0 study. We evaluated the results and challenges of the first pilot. The study protocol was revised to include more participants and the use of new smart technology, specifically the Corsano CardioWatch, replacing the non-CE marked devices from the first pilot, which required 24/7 researcher availability. Preparations included finalizing the application to the Medical Ethical Review Committee and thinking out the study logistics, resulting in a comprehensive plan for the conduct of the study. Key takeaways from this process include the recognition that research is an iterative process of continuous learning, that the purpose of the study must be carefully considered with ethical concerns in mind, and that the increasing importance of data requires careful planning for its security, processing and storage. The METC application has been submitted. Once approved, the study can proceed on a well-prepared basis.

# Table of Contents

# List of Abbreviations

| | |
|---|---|
| **AI** | Artificial intelligence |
| **ANS** | Autonomic nervous system |
| **APACHE** | Acute physiology and chronic health evaluation |
| **AUC** | Area under the curve |
| **DWH** | Data warehouse |
| **ECG** | Electrocardiogram |
| **FFT** | Fast Fourier transform |
| **HR** | Heart rate |
| **HF** | High frequency |
| **HRV** | Heart rate variability |
| **Hz** | Hertz |
| **ICU** | Intensive care unit |
| **IQR** | Interquartile range |
| **LF** | Low frequency |
| **LOS** | Length of stay |
| **LUMC** | Leiden University Medical Center |
| **MAE** | Mean absolute error |
| **MDR** | Medical device regulation |
| **METC** | Medical ethical review commission |
| **PICS** | Post intensive care syndrome |
| **PNS** | Parasympathetic nervous system |
| **PSD** | Power spectral density |
| **RMSSD** | Root mean square of successive differences between normal to normal RR intervals |
| **ROC** | Receiver operating curve |
| **SD** | Standard deviation |
| **SDANN** | Standard deviation of 5-minute average normal to normal RR intervals |
| **SDNN** | Standard deviation of all normal to normal RR intervals |
| **SHAP** | Shapley additive explanations |
| **SNS** | Sympathetic nervous system |
| **SOFA** | Sequential organ failure assessment |
| **SQL** | Structured query language |
| **SVM** | Support vector machine |
| **SVR** | Support vector regressor |
| **ULF** | Ultra-low frequency |
| **VLF** | Very low frequency |
| **WMO** | Medical research involving human subjects act |
| **XGB** | Extreme gradient boosting |

# List of Figures

# List of Tables

# 1

# Improved Intensive Care Unit outcome prediction based on Heart Rate Variability


## Abstract

**Background:** The ICU is critical for patients who require intensive monitoring or organ support. ICU admissions are associated with high mortality and morbidity, making it essential to identify patients at increased risk of complications. HRV, the fluctuation in time intervals between consecutive heartbeats, shows promise as a prognostic marker.
**Objective:** The aim of this study was to determine whether heart rate variability is predictive of ICU length of stay and ICU mortality using a machine learning approach.
**Methods:** We retrospectively analyzed data from all adult patients admitted to the ICU of the LUMC for at least 48 hours between February 2023 and March 2024. We derived nine HRV measures from electrocardiograms and extracted additional clinical features from electronic health records. We then evaluated the performance of three classifiers for predicting ICU mortality and three regressors for predicting ICU length of stay.
**Results:** We included 468 patients of which 71 died during ICU admission. The eXtreme Gradient Boosting model, incorporating clinical features and three short-term HRV measures (i.e., the power in the high frequency band, the power in the very low frequency band and the low frequency/high frequency ratio) achieved the highest area under the curve of 0.76 for predicting ICU mortality. Predictive models for ICU length of stay were less effective, with the best model achieving a mean absolute error of 5.07 days.
**Conclusion:** HRV, when combined with clinical features, provides valuable predictive information for ICU mortality. Models using clinical features with or without HRV measures did not accurately predict ICU length of stay. Further research is needed to fully explore the potential of HRV to predict both ICU mortality and length of stay before considering its clinical implementation.

## 1.1 Introduction

An intensive care unit (ICU) is a specialized hospital setting providing intensive and specialised medical and nursing care to critically ill patients, including extensive monitoring and various forms of organ support to sustain life during acute organ system insufficiency (1). An ICU admission is not only essential for critically ill patients but is also often required postoperatively for patients who have undergone major surgery such as cardiothoracic surgery. According to the European Society of Intensive Care Medicine, per year, about five million adults are admitted to ICUs worldwide (2). Even though an increasing number of critically ill patients survive the ICU and are discharged to their homes due to advances in technology and practice, the ICU remains associated with both high mortality and morbidity (3–5). A recent study into the survival of ICU patients showed an overall ICU mortality of 16% and an in-hospital mortality of 24% (6).

It is crucial to identify patients who face an elevated risk of complications during ICU admission. In recent years, there has been an increasing focus on predictive models due to advances in technology and machine learning (7). Nevertheless, making accurate predictions with the data available remains challenging, even, today, for conventional vital signs.

However, a non-conventional parameter that appears to have promising prognostic value but requires further research is heart rate variability (HRV) (8). HRV is the fluctuation in the time intervals between consecutive heartbeats and a measure of autonomic nervous system (ANS) activity on the heart (9). In healthy individuals, there is a dynamic balance between the two arms of the ANS, the parasympathetic nervous system (PNS) and the sympathetic nervous system (SNS). This balance can be disturbed in critical illnesses and is often seen in critically ill patients in the ICU (10). Evaluating this aspect of autonomic function in critically ill patients can offer insight into pathophysiology, disease severity, response to treatment and prognosis.

## 1.2 Background

### 1.2.1 Heart rate variability

Heart rate variability is a term to describe the fluctuation in the time intervals between consecutive heartbeats (9). We can derive these intervals, known as RR intervals, from an electrocardiogram (ECG) (Figure 1.1).

The ANS, a component of the peripheral nervous system that regulates involuntary physiological processes, constantly alters and regulates heart rate and rhythm (11). The PNS and SNS control heart rate through different mechanisms (12). The PNS slows heart rate by releasing acetylcholine, while the SNS increases heart rate and contractility by releasing catecholamines. These effects occur at different speeds, with the PNS acting almost immediately and the SNS having delayed onset but longer-lasting effects. The PNS regulates heart rate on a beat-by-beat basis, while the SNS affects heart rate for several seconds after stimulation ends (5-10 seconds).

HRV can be captured by a variety of measures that are calculated in either the time or frequency domain. Time domain measurements of HRV involve statistical calculations, such as the standard deviation and the root mean square of successive differences of RR intervals, which quantify the variability in the time intervals between consecutive heartbeats (9).

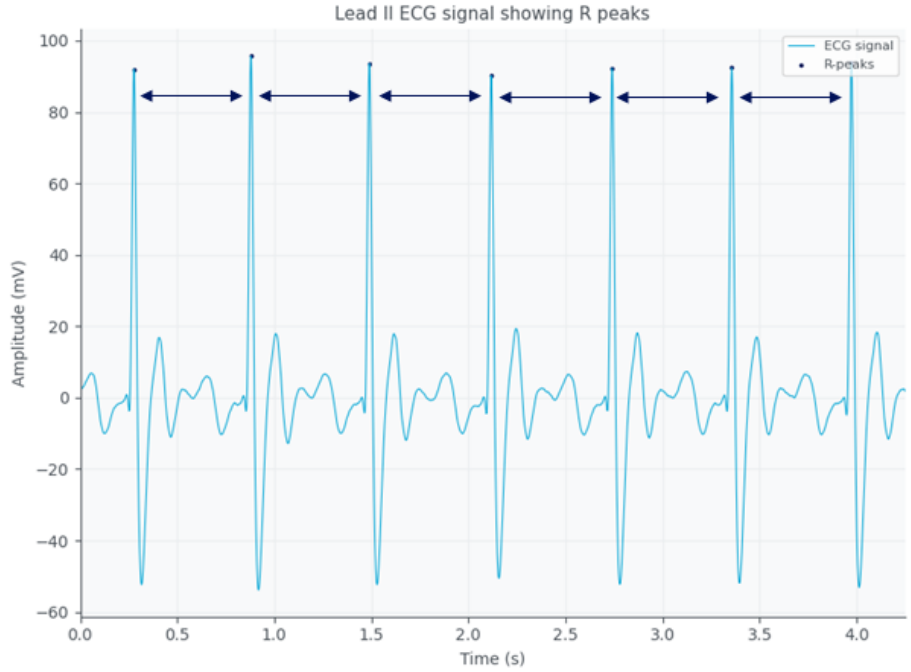



**Figure 1.1: The RR intervals of the electrocardiogram signal.**

Frequency domain analysis, which uses Power Spectral Density (PSD), employs a Fast Fourier Transform (FFT) to convert fluctuations in RR intervals into specific frequency ranges (12). This helps to understand the physiological mechanisms behind HRV. Spectral analysis shows how power is distributed across different frequencies of rhythms in RR-fluctuations (13). It measures how strong (amplitude) and how often (frequency) these rhythms occur. The results are expressed as power spectral density, which is the area under the peak in a frequency range. A visual representation of the steps from RR intervals to frequency domain parameters is shown in Figure A.1 of Appendix A.1. The Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology divided RR interval oscillations into ultra-low frequency (ULF), very low frequency (VLF), low frequency (LF) and high frequency (HF) bands (14).

HRV can be recorded over different periods of time, typically categorized as short-term measurements of 5 minutes or long-term measurements of 24 hours. Each combination of HRV measure and recording duration provides a reflection of the ANS. The autonomic, cardiovascular, central nervous, endocrine, and respiratory systems, alongside baroreceptors and chemoreceptors, impact HRV over a short time period, contributing to the very low to high frequencies of the spectrum (6). Factors such as circadian rhythms, core body temperature, metabolism, sleep cycles, and the renin-angiotensin system are believed to contribute to HRV recordings of 24 hours. Therefore, short-term values cannot be substituted for 24-hour values, even though the mathematical formula may be the same.

Unlike conventional vital signs, which also provide insight into a patient's health status, HRV serves as a comprehensible and non-invasive measure of autonomic dysregulation. HRV has been shown to have potential as a useful predictor of several (critical) illnesses, such as neurological disorders, cardiovascular disorders, infection, sepsis, septic shock, multiple organ dysfunction and severe trauma (10). It has also proven to be a non-specific predictor of mortality (16). There is no consensus in the literature as to whether increased or decreased HRV leads to poorer outcomes, with studies suggesting that both higher and lower HRV can be associated with poor outcomes (6).

### 1.2.2 Machine Learning

Machine learning is a form of artificial intelligence (AI) that uses mathematical and computational systems to extract information from data, typically for the purpose of prediction (15). One subtype of machine learning is supervised machine learning, an approach that generalizes information from a training set's features to create a model that can correctly predict outcomes. Subsequently, this learned model is applied to make predictions using the unseen features from a testing data set (16). Advantages of (supervised) machine learning compared to conventional statistical models are its ability to incorporate a larger number of variables, allowing it to make more powerful predictions, and its ability to identify trends or patterns in large datasets that might be missed by researchers (17). Nowadays, machine learning is widely adapted in intensive care medicine. For instance, studies have demonstrated its effectiveness in predicting clinical outcomes such as the risk for ICU transfer, cardiac arrest, or mortality in ICU patients (18).

## 1.3 Methods

### 1.3.1 Objective

The objective was to determine if heart rate variability is predictive of clinical outcomes in ICU patients, i.e., ICU length of stay (LOS) and ICU mortality. To achieve this objective, we extracted the HRV measures from ECG data while also incorporating additional patient data to develop other clinical features. We then applied machine learning techniques to assess the predictive value, as shown in Figure 1.2.

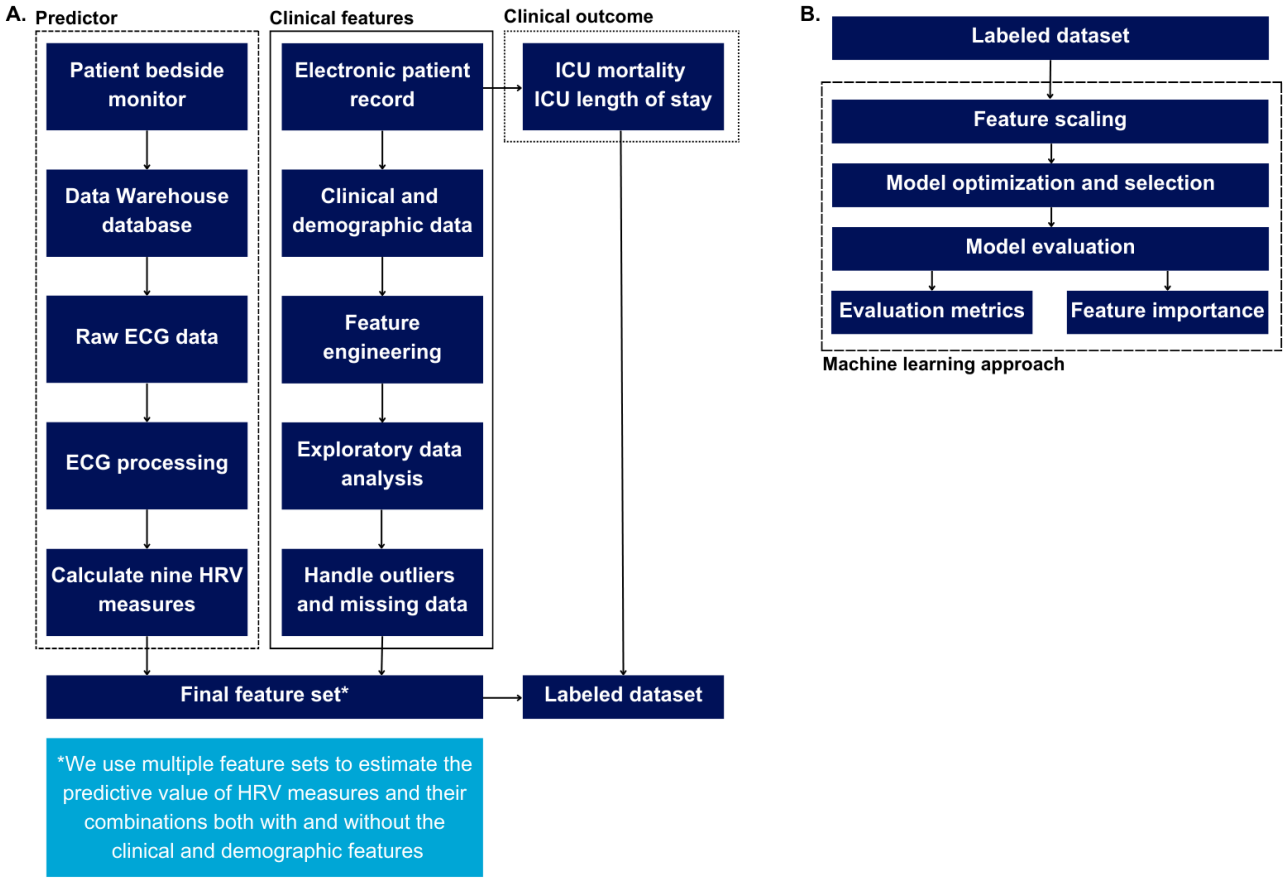



**Figure 1.2: A schematic overview of the Method.** A. shows the preprocessing of the raw data and the creation of the labeled dataset and B. shows the machine learning approach.

### 1.3.2 Study design

This was a single-centre, retrospective, prognostic study.

### 1.3.3 Subjects

All patients admitted to the ICU of the Leiden University Medical Center (LUMC) between March 1st, 2023 and February 29th, 2024 were retrospectively reviewed. We included adult patients (18 years or older) admitted for 48 hours or longer. This study was not subject to the Medical Research Involving Human Subjects Act (WMO). We received approval from the local ethics review committee and the requirement to obtain informed consent was waived.

### 1.3.4 Data availability

We used two different types of data: high-frequency data obtained from the patient's bedside monitor and data extracted from the electronic health record. The bedside monitor displays various clinical values, consisting of waveform and numerical data, and stores them in a database referred to as Data Warehouse (DWH). We sampled waveform data, like ECG data, at 500 Hz. Numeric data, such as heart rate and blood pressure, were sampled at 1 Hz. All DWH data were available for the entire ICU admission; however, we used only the data from the first 24 hours of admission. There may have been brief interruptions in the data due to the patient undergoing (re)procedures or diagnostic tests at another location within the hospital. There was no fixed sampling frequency for data from the electronic health record. This data included one-time recordings upon admission or discharge, such as gender or admission diagnosis, or at fixed intervals during each shift, such as clinical (risk) scores. Data such as medication administration and lab values were always documented at the time of the event, but the specific details and frequency of these events varied depending on the individual patient.

The DWH data and the electronic health record data were linked by patient number. Each participant's study ID was linked to the patient numbers. In the final datasets, this patient number was removed. It was stored in a key file on a secure server that was only accessible to the researchers. Although it no longer contained any traceable patient information, the rest of the dataset was also stored on a secure LUMC server.

### 1.3.5 Electrocardiogram processing

From the DWH data, lead II of the ECG was available for each patient and was used to calculate the HRV measures. The DWH database stored the ECG data in hexadecimal format. Using Structured Query Language (SQL), the raw data were imported into Python for further processing. A schematic overview of all the steps involved in ECG processing is shown in Figure 1.3. A visual step-by-step overview showing how the ECG signal was affected at each step is provided in Figure A.2 in Appendix A.2.

We started by converting the hexadecimal data into decimal values. The function clean_ecg from Python package neurokit2 was used to apply the Pan-Tompkins method to the raw ECG data (19). The Pan-Tompkins method involves applying a band-pass filter to the signal, followed by differentiation, squaring, and moving window integration (20).

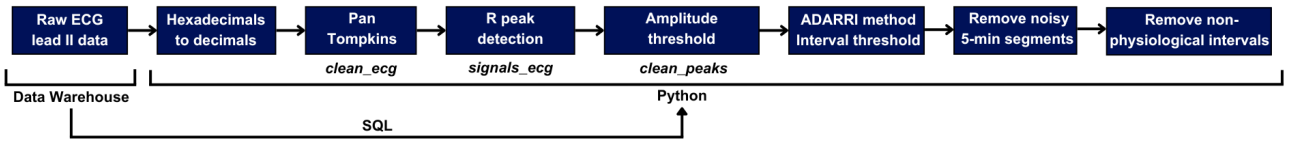



**Figure 1.3: A schematic overview of all steps involved in the processing of the ECG and the RR intervals.**

R-peaks were then detected using the function signals_ecg from Python package biosppy (21). To ensure accurate HRV calculation of only the normal RR intervals, falsely detected R-peaks caused by movement artifacts or R-peaks from ectopic beats needed to be removed. Ectopic beats are heartbeats that do not originate from the sinus node are often premature, and therefore do not directly reflect the activity of the autonomic nervous system (22). First, the function clean_peaks was applied to the initially detected R-peaks. This function was based on a method used by Van Wijk et al. in a study where they calculated HRV measures using ECG recordings (23). This function calculated the mean amplitude of the R-peaks over a five-minute period, then set a minimum threshold of one-third of the mean and a maximum threshold of the mean times 1.6 for each peak.

Then, we calculated the RR intervals based on the remaining R-peaks. The clean_peaks function removed R-peaks outside the limits set by the thresholds, resulting in RR intervals that were too large. Because of this and the presence of ectopic beats, there were both excessively large and small RR intervals. These outlier intervals were removed using the ADARRI method (24). The ADARRI method uses the absolute difference between consecutive RR intervals, deeming an adjacent RR interval non-physiological if it exceeds the optimal threshold of 276 ms.

Hereafter, we implemented a quality control step to identify and remove five-minute segments heavily affected by artifacts or ectopic beats. This was achieved by calculating the percentage of initially detected R-peaks that were subsequently removed during clean_peaks and ADARRI steps. Five-minute segments exceeding a 35% removal threshold were excluded from HRV calculations. We established this threshold through manual trial and error based on visual inspection, balancing the removal of excessive data with minimizing the inclusion of noisy data.

The final step involved removing all RR intervals outside the range of 333 to 1500 milliseconds, corresponding to a minimum heart rate of approximately 40 beats per minute and a maximum heart rate of approximately 180 beats per minute. These boundaries were selected in close consultation with an intensivist. For each patient, a visual inspection was performed to determine whether the full ECG processing algorithm was performing correctly, and patients with ECGs that contained too much noise or morphology incompatible with the ECG processing steps were excluded.

### 1.3.6 Heart rate variability calculation

After processing the ECG, we calculated HRV measures from the remaining RR intervals using the Python package pyhrv (25). The Task Force created guidelines for standards of measurement for HRV (14). It is recommended to use 5-minute/short-term recordings for four frequency domain measures: the power in the VLF, LF and HF band and the LF/HF ratio. For 24-hour or long-term measurement, it is recommended to use four time domain measures and one frequency domain measure: the root mean square of successive differences of the normal RR intervals (RMSSD), the standard deviation of the normal RR intervals (SDNN), the standard deviation of the average normal

RR intervals (SDANN), the triangular index and the power in the ULF band. More recently, it was discovered that RMSSD could be accurately calculated over a 5-minute period (9). This recording duration is now the gold standard for RMSSD measurement, which is why we included RMSSD in the short-term measures. In addition to the Task Force recommendations, we conducted a brief but comprehensive review of studies that also investigated the predictive value of HRV in a critically ill population (23, 26–36). A brief overview of these studies is provided in Table A.1 of Appendix A.3. In addition to the measures recommended by the Task Force, a few additional HRV measures were used in these studies. However, the most frequently used HRV measures were the nine that were recommended by the Task Force, so we opted to use these as the primary metrics for our study (Table 1.1).

**Table 1.1: The included HRV measures and their reflection of the ANS.** (14, 37). HR: heart rate, HRV: Heart Rate Variability, Hz: hertz, ms: milliseconds, PNS: parasympathetic nervous system, SNS: sympathetic nervous system.

| Measure | Domain | Meaning | Reflection of the autonomic nervous sytem |
|---|---|---|---|
| *Short term* | | | |
| RMSSD (ms) | Time | The root mean square of successive differences between normal RR intervals | Reflects parasympathetic activity |
| VLF ($ms^2$) | Frequency | Power in the very low frequency range 0.003 - 0.04 Hz | Reflects regulation mechanisms, thermoregulation and hormonal mechanisms. |
| LF ($ms^2$) | Frequency | Power in low frequency range 0.04 - 0.15 Hz | Reflects a mix of sympathetic, parasympathetic activity and baroreflex activity. |
| HF ($ms^2$) | Frequency | Power in high frequency range 0.15 - 0.4 Hz | Reflects parasympathetic activity and is linked to heart rate changes associated with the respiratory cycle. |
| LF/HF | Frequency | Ratio LF/HF | Reflects a mix of sympathetic and vagal activity. |
| *Long term* | | | |
| SDNN (ms) | Time | Standard deviation of all normal RR intervals | Reflects all the cyclic components responsible for variability, both sympathetic and parasympathetic activity. |
| SDANN (ms) | Time | Standard deviation of 5-minute average normal RR intervals | Is similar to SDNN, but minimizes the effects of editing, artefacts and missed or ectopic beats. |
| HRV triangular index | Time | Total number of all normal RR intervals divided by the height of the histogram of RR intervals measured on a discrete scale with bins of 7.8125 ms (1/128 seconds) | Reflects overall HRV measured and is more influenced by the lower than by the higher frequencies. Permits for only casual preprocessing of the ECG signal. |
| ULF ($ms^2$) | Frequency | Power in the ultra-low frequency range ≤0.003 Hz | Reflects circadian oscillations, core body temperature, metabolism and the renin-angiotensin system. |

The five-minute segment with the lowest percentage of removed R-peaks in the first hour of admission was used to calculate the short-term HRV measures to ensure the most reliable R-peaks were used. As for the long-term measures, the values were calculated over the first 8 hours of admission. Due to computational issues, the long-term measures could not be calculated over a 24-hour period.

### 1.3.7 Clinical features

In addition to the HRV measures that served as the predictor of interest, we included a total of 43 demographic and clinical features to improve the accuracy and robustness of the model. The additional features, which are known to be related to ICU mortality and ICU LOS, allowed the identification of complex interactions between HRV and other patient and ICU-related characteristics. To ensure clinical relevance, these features were carefully selected in consultation with an intensivist. The specifics of the features, including their types and the times at which they were measured, are shown in Table A.2 of Appendix A.4.

### 1.3.8 Clinical feature engineering

Three features had to be adjusted before they could be used in the model. Gender was converted from male/female to a binary variable where male = 1 and female = 0. The admission type feature consisted of three categories: medical admission, emergency surgery and planned surgery. One hot encoding was used to adjust this feature (38). The acute physiology and chronic health evaluation (APACHE) IV admission diagnosis category consisted of 145 different categories denoted by integers. A drawback of using a single feature with 145 unique integer values is that the machine learning algorithm may misinterpret these integers as having an ordinal relationship (39). Besides, if we were to apply one hot encoding to this feature, 144 new columns would have to be added, resulting in a huge increase in the size of the dataset, which would slow down the learning of the model and degrade the overall performance (40). Instead of having 145 integers representing the diagnosis categories, we replaced these with the mortality coefficients of the diagnoses, which were available from the Nationale Intensive Care Evaluatie (NICE) foundation dictionary.

An exploratory data analysis was performed to identify outliers and missing data. Missing numerical data were imputed using the median, while missing categorical data were replaced with the most frequent value. Outliers were manually reviewed, assessed and acted upon as appropriate. The treatment of outliers is described in more detail in Appendix A.5.

### 1.3.9 Intensive care unit outcomes

The clinical outcomes ICU LOS and ICU mortality were available from the electronic health record. ICU LOS was expressed in days of 24 hours and ICU mortality was defined as 1 = non-ICU survivor and 0 = ICU-survivor.

### 1.3.10 Model optimization and selection

The ICU length of stay is a continuous outcome, whereas ICU mortality is a categorical outcome. We used regression models to predict ICU length of stay, whereas the models predicting ICU mortality used a classification approach. We evaluated a Lasso regression model, a support vector regression (SVR) model, and an Extreme Gradient Boosting (XGBoost) model for the ICU LOS prediction model. We evaluated a logistic regression model, a support vector machine (SVM) model, and an XGBoost model for the ICU mortality prediction model. We defined a set of possible hyperparameter settings for each model, which are shown in Table A.3 in Appendix A.6. We created a pipeline for each model to ensure a standardized and reproducible workflow. This approach included scaling non-binary features using the StandardScaler function from the Python scikit-learn package (42). Standardization is a common requirement for many machine learning models to perform optimally, especially when features deviate from a normal distribution.

To optimise and select the final model, we used nested cross-validation (Figure 1.4). We started by performing an outer cross-validation across the entire dataset using 10 folds with stratified splits. Within each fold of the outer cross-validation, an inner cross-validation with 5 folds was conducted, using stratified splits as well. This inner cross-validation served the purpose of hyperparameter optimization and model selection. From the five folds of the inner cross-validation, the model with the highest performance was selected as the best model of the corresponding outer cross-validation fold. Subsequently, this selected model was evaluated on the test set of this fold.

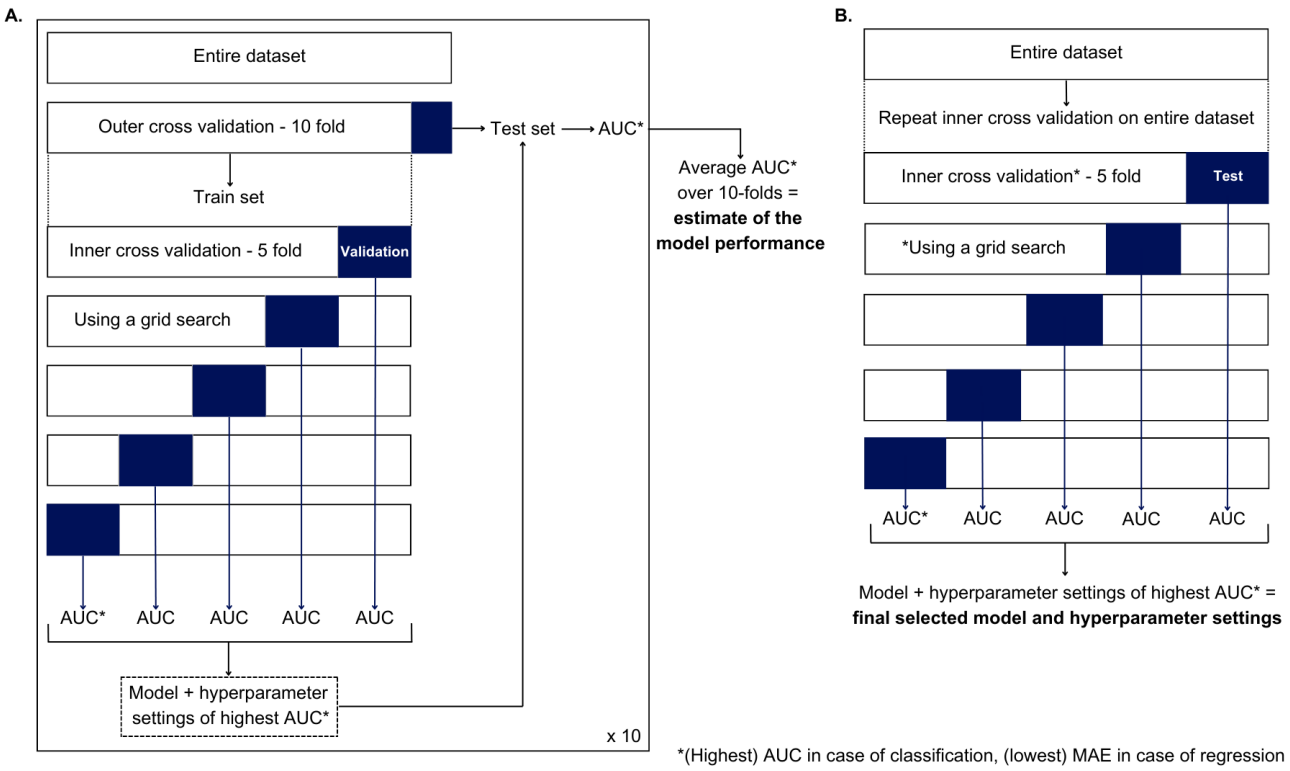



**Figure 1.4: A schematic overview of the model optimization and selection using nested cross validation.** A. shows an overview of getting an estimate of the model's performance and B. shows the model selection process.

For the regression problem, we used the Mean Absolute Error (MAE) as a performance metric and for the classification problem, we used the Area Under the Curve (AUC). The 10-fold outer cross-validation yielded a maximum of 10 different models, one for each fold, each associated with its respective MAE/AUC score. The average performance score across these 10 models was computed and served as an estimate of the model performance. This was the most reliable estimate because it reflected how the model performed on unseen data through multiple folds.

Following this evaluation, we configured the final model by applying the inner cross-validation to the entire dataset to select the best model based on performance across the entire dataset.

### 1.3.11 Model evaluation

The mean MAE and AUC from the outer 10-fold cross-validation served as the primary metrics for evaluating model performance. In addition to these metrics, we translated the model's performance into other metrics. These additional metrics were also calculated during the 10-fold cross-validation by averaging the results from all folds. For the ICU LOS prediction, the other measures included the root mean squared error (RMSE) and the R squared ($R^2$) metric. For the ICU mortality prediction, the additional measures were sensitivity, precision and the F1-score. In the results section, mentions of feature set performance (of all evaluation metrics) refer to the average of the 10 outer folds unless otherwise stated.

In addition to identifying which HRV measure(s) would be the most predictive, we also wanted to determine the extent to which the HRV measure contributed to that outcome, especially when combined with other clinical features. Therefore, we determined feature importance scores and SHapley

Additive exPlanations (SHAP) values for the model with the highest performance.

### 1.3.12 Analysis plan

We aimed to evaluate the predictive value of the individual HRV measures, the individual HRV measures combined with clinical features, and combinations of HRV measures with clinical features.

We started with the incorporation of one of the selected HRV measures, allowing us to evaluate the performance of each measure individually, without the context of the other selected clinical features. We then examined the performance of the clinical features alone to establish baseline performance without any HRV measures. Next, we evaluated the effect of all clinical features combined with each individual HRV measure. We then wanted to explore the performance of all HRV measures in pairs of two and combinations of multiple HRV measures, i.e. all short-term measures combined, all long-term measures combined and all HRV measures combined. However, correlations between various HRV measures have previously been described in the literature (6, 41–43). Multicollinearity can lead to poor generalization and overfitting of the data, negatively affecting model performance on unseen data (44). Therefore, before evaluating the predictive models using combinations of HRV measures, we created a correlation matrix to gain insight into the interactions between HRV measures, allowing us to make informed decisions to mitigate multicollinearity, such as removing or combining highly correlated measures.

We decided to evaluate all pairwise combinations of HRV measures to provide a complete overview, regardless of the correlations. For combinations including multiple HRV measures (> 2), we assessed correlations using the rule of thumb for interpreting the size of a correlation coefficient (45). We refrained from combining HRV measures with high and very high correlations (> 0.70) and also aimed to minimize the use of HRV combinations with moderate correlations (0.50 - 0.70).

### 1.3.13 Statistics

Descriptive data were reported as mean ± SD for normally distributed continuous variables or median with interquartile range [IQR] for non-normally distributed variables, unless stated otherwise. Categorical variables, such as gender, comorbidities, and admission type, were expressed as percentages. Normality was tested using the Kolmogorov-Smirnov test. Group differences in dichotomous variables were tested using Fisher's exact test, while group differences in continuous data were tested using the Mann-Whitney U test. P-values were considered significant at $p < 0.05$.

## 1.4 Results

### 1.4.1 Population

Between March 1st, 2023 and February 29th, 2024, 517 patients with 581 admissions were admitted to the ICU of the LUMC with a length of stay of at least 48 hours. Among these patients, 468 admissions from 453 patients were included in our study. Eighty-three admissions were excluded due to discrepancies between the DWH database and the electronic patient record. Twenty-nine admissions were excluded based on visual inspection of the ECG. Either the signals contained too much noise, resulting in no usable data, or the ECG processing algorithm failed to process that specific ECG. Each admission was treated as a separate patient due to potential differences in baseline

and disease progression over time. Therefore, we referred to 468 patients and based all calculations on this number.

The median [IQR] age was 62 [52 - 71] and 65% were male. The median [IQR] APACHE IV score was 66 [51 - 84]. 59% had a medical admission, 23% were admitted after planned surgery and 18% were admitted after emergency surgery. Within the first 24 hours of admission, 68% of patients required mechanical ventilation. The median [IQR] LOS was 5 [3 - 10] days. Seventy-one patients died during ICU admission. Complete demographic and clinical information is listed in Table 1.2 and characteristics of the ICU admission are listed in Table 1.3.

**Table 1.2: Demographic and clinical characteristics of the included patients.** The p-value indicates the statistical significance of the differences between the groups, as determined by Fisher's exact test (for dichotomous data) or the Mann-Whitney U test (for continuous data). A p-value less than 0.05 was considered significant.

| Characteristic | All n = 468 | ICU survivors n = 398 | ICU non-survivors n =71 | p-value |
|---|---|---|---|---|
| Age (years) | 62 [52-71] | 62 [52-71] | 65 [56-71] | 0.303 |
| Gender (% male) | 64.7% | 64.2% | 67.7% | 0.686 |
| BMI (kg/m²) | 26.0 [23.0 - 29.0] | 26.0 [23.1 - 28.8] | 26.1 [22.9 - 31.1] | 0.316 |
| APACHE IV | 66.0 [51.0 - 83.8] | 63.0 [49.0 - 80.0] | 84.0 [68.0 - 98.0] | **<0.001** |
| Highest SOFA during first 24 hours of admission | 8 [6 - 10] | 8 [6 - 9] | 9 [8 - 13] | **<0.001** |
| Medical admission (%) | 59.2% | 56.2% | 76% | **0.002** |
| Planned surgery (%) | 23.3% | 24.9% | 14.1% | **0.048** |
| Emergency surgery | 17.5% | 18.9% | 9.9% | 0.088 |
| *Admission diagnosis* | | | | |
| Cardiovascular | 49.6% | 49.1% | 52.1% | 0.700 |
| Gastrointestinal | 11.8% | 11.6% | 12.7% | 0.841 |
| Genitourinary | 1.7% | 2.0% | 0% | 0.614 |
| Hematology | 0.4% | 0.5% | 0% | 1.000 |
| Metabolic/Endocrine | 0.6% | 0.8% | 0% | 1.000 |
| Musculoskeletal/skin | 0.9% | 0.8% | 1.4% | 0.483 |
| Neurologic | 8.1% | 8.6% | 5.6% | 0.488 |
| Respiratory | 17.5% | 17.6% | 16.9% | 1.000 |
| Transplant | 3.2% | 3.3% | 2.8% | 1.000 |
| Trauma | 6.2% | 5.8% | 8.5% | 0.420 |
| *% of patients with comorbidities at admission* | | | | |
| Acquired immune deficiency syndrome | 0.0% | 0.0% | 0.0% | - |
| Cardiopulmonary resuscitation | 9.9% | 9.1% | 14.3% | 0.192 |
| Chronic dialysis | 0.9% | 1% | 0% | 1.000 |
| Chronic cardiovascular insufficiency | 1.1% | 1% | 1.4% | 0.563 |
| Chronic obstructive pulmonary disease | 3.2% | 3.1% | 4.2% | 0.486 |
| Chronic renal insufficiency | 7.1% | 6.6% | 10% | 0.312 |
| Chronic respiratory insufficiency | 2.6% | 2.0% | 5.6% | 0.093 |
| Cirrhosis | 3.4% | 3.3% | 4.2% | 0.721 |
| Diabetes | 10.1% | 9.6% | 12.7% | 0.397 |
| Hematologic malignancy | 3.4% | 2.5% | 8.5% | **0.023** |
| Immunological insufficiency | 4.3% | 4.0% | 5.6% | 0.525 |
| Metastatic neoplasm | 0.9% | 1% | 0% | 1.000 |

**Table 1.3: Intensive care unit treatment characteristics.** The p-value indicates the statistical significance of the differences between the groups, as determined by Fisher's exact test (for dichotomous data) or the Mann-Whitney U test (for continuous data). A p-value less than 0.05 was considered significant.

| Characteristic | All n = 468 | ICU survivors n = 398 | ICU non-survivors n = 71 | p-value |
|---|---|---|---|---|
| *Vital parameters of the first 24 hours of admission (mean ± SD)* | | | | |
| Diastolic blood pressure | 58.9 ± 8.6 | 58.9 ± 8.8 | 58.4 ± 8.6 | 0.836 |
| Heart frequency | 87.2 ± 19.5 | 86.0 ± 18.8 | 93.4 ± 22.4 | **0.014** |
| Mean arterial blood pressure | 76.8 ± 8.9 | 76.9 ± 9.0 | 76.2 ± 7.9 | 0.514 |
| Oxygen saturation | 96.4 ± 1.9 | 96.5 ± 1.8 | 95.8 ± 1.9 | **0.020** |
| Respiratory rate | 17.93 ± 4.17 | 17.74 ± 4.14 | 19.02 ± 4.19 | **0.010** |
| Systolic blood pressure | 114.97 ± 16.47 | 115.31 ± 16.41 | 113.12 ± 16.83 | 0.131 |
| *% of patients with comorbidities after first 24 hours of admission* | | | | |
| Acute renal failure | 13.5% | 10.1% | 32.4% | **<0.001** |
| Mechanical ventilation | 68.1% | 67.2% | 73.2% | 0.336 |
| *% of patients who received medication within the first 24 hours of admission* | | | | |
| Adrenalin | 3% | 2.3% | 7.6% | **0.046** |
| Amiodaron | 7.7% | 6.8% | 12.7% | 0.093 |
| Clonidine | 9.8% | 10.6% | 5.6% | 0.278 |
| Dexmedetomidine | 0.6% | 0.8% | 0% | 1.000 |
| Dobutamine | 12.2% | 10.8% | 19.7% | **0.047** |
| Bumetanide | 1.1% | 1.0% | 1.4% | 0.562 |
| Furosemide | 17.5% | 18.4% | 12.7% | 0.309 |
| Labetalol | 5.3% | 6.3% | 0% | **0.021** |
| Metoprolol | 3.6% | 4% | 1.4% | 0.490 |
| Sotalol | 1.9% | 2.0% | 1.4% | 1.000 |
| Midazolam | 10.3% | 8.3% | 21.1% | **0.002** |
| Nitroglycerine | 2.1% | 2.0% | 2.8% | 0.653 |
| Nitroprusside | 5.1% | 6.1% | 0% | **0.036** |
| Noradrenaline | 73.1% | 70.1% | 90.1% | **<0.001** |
| Propofol | 58.3% | 56.7% | 67.7% | 0.091 |
| Sufentanil | 48.9% | 46.4% | 63.4% | **0.010** |
| *Lab values* | | | | |
| Highest C-reactive protein during first 24 hours of admission | 60.6 [19.6 – 150.2] | 57.7 [17.2 – 140.2] | 98.2 [32.6 – 243.5] | **0.015** |
| *Length of stay* | | | | |
| Length of stay (days) | 4.9 [3.0 - 9.6] | 4.8 [3.0 – 8.9] | 7.4 [4.0 – 17.1] | **<0.001** |

### 1.4.2 Heart rate variability measures

We calculated five short-term HRV measures and four long-term HRV measures (Table 1.1). The mean and standard deviation of each measure are shown in Table 1.4 for the entire population and for ICU survivors and ICU non-survivors separately. Two HRV measures, the power in the HF band and the SDANN, showed a significant difference between the two groups, where the power in the HF band was higher and the SDANN was lower in non-survivors.

To gain insight into the interactions between the HRV measures, we explored their correlation (Figure 1.5). We observed high correlations (>0.80) between 1) the RMSSD and the power in the HF band, 2) the power in the LF band and the power in the HF band and 3) the SDNN and SDANN. As the power in the LF band and the RMSSD were also somewhat correlated (0.68), we decided to discard these two and include the power in the HF band when combining multiple HRV measures, especially as the power in the HF band demonstrated a significant difference between ICU survivors and non-survivors. In addition, the SDNN showed some correlation with the triangular index (0.57) while the SDANN did not. Given that we wanted to minimize correlations, we decided to discard the

SDNN and include the SDANN when combining multiple HRV measures, also because the SDANN showed a significant difference between ICU survivors and non-survivors as well.

**Table 1.4: Values of the nine included heart rate variability measures.** Data are presented as mean ± SD. The p-value indicates the statistical significance of the differences between the groups, tested using the Mann-Whitney U test. A p-value <0.05 was considered significant.

| HRV measure | All | ICU survivors | ICU non-survivors | p-value |
|---|---|---|---|---|
| *Short-term* | | | | |
| RMSSD (ms) | 68.0 ± 27.0 | 66.8 ± 26.2 | 74.7 ± 30.8 | 0.071 |
| VLF (ms²) | 307.3 ± 864.5 | 296.9 ± 851.6 | 365.2 ± 937.7 | 0.257 |
| LF (ms²) | 385.5 ± 821.0 | 357.0 ± 764.1 | 544.9 ± 1078.6 | 0.230 |
| HF (ms²) | 1202.0 ± 1341.8 | 1155.0 ± 1354.9 | 1465.1 ± 1242.4 | **0.013** |
| HF/LF ratio | 0.23 ± 0.23 | 0.22 ± 0.22 | 0.24 ± 0.24 | 0.702 |
| *Long-term* | | | | |
| SDNN (ms) | 85.9 ± 34.9 | 86.6 ± 35.3 | 82.1 ± 32.5 | 0.290 |
| SDANN (ms) | 57.4 ± 38.0 | 58.7 ± 38.4 | 50.3 ± 35.1 | **0.028** |
| Triangular index | 12.9 ± 7.7 | 13.1 ± 7.7 | 11.8 ± 7.6 | 0.109 |
| ULF (ms²) | 15.6 ± 45.8 | 16.3 ± 49.3 | 11.9 ± 15.8 | 0.221 |

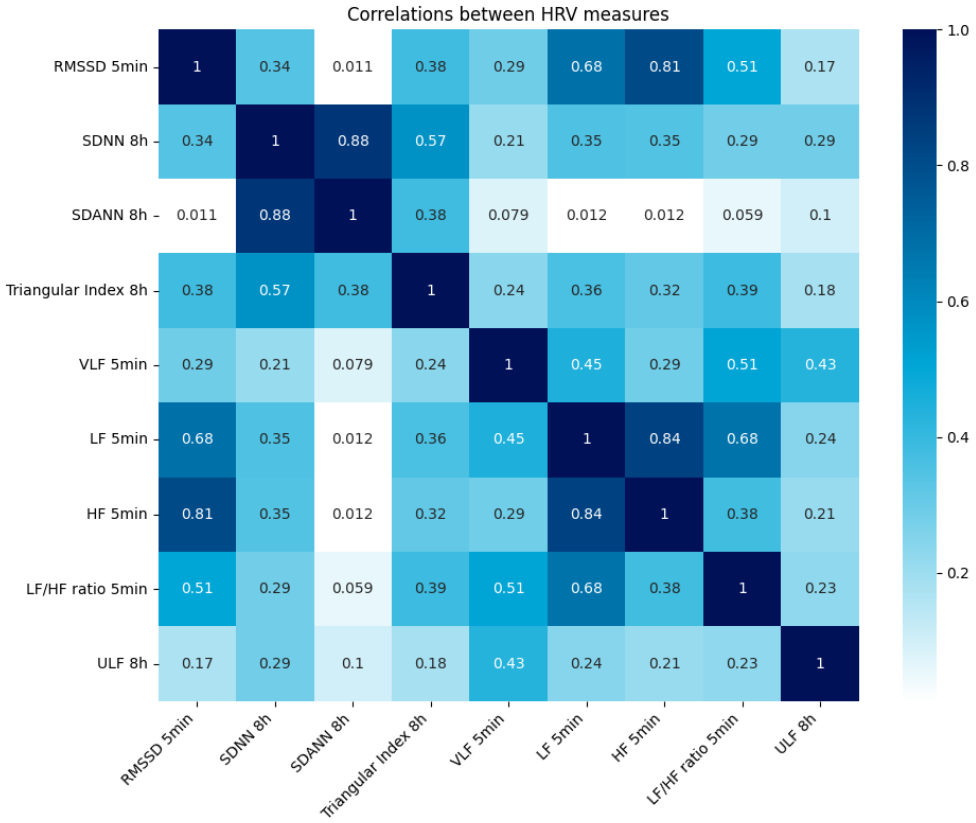



**Figure 1.5: Heatpmap showing the Pearson correlation between the nine HRV measures.** The color scale indicates the strength of the correlations, with high correlations in red and low correlations in red.

### 1.4.3 Predictive models for intensive care unit mortality

First, we tested predictive models for ICU mortality for individual HRV measures without including any other relevant clinical features. We explored three different models: a Logistic Regression model, a SVM model and an XGBoost classifier. We used nested-cross validation for model selection and performance estimation (Figure 1.4). As described in the Methods Section, all mentions

of feature set performance refer to the average of the 10 outer folds of the nested cross-validation, unless otherwise stated.

All AUCs were low, ranging from 0.46 to 0.57 for the best model choice (Figure 1.6). The highest AUC was achieved by an SVM classifier using the RMSSD measure. Sensitivity, precision and F1 score were all very low, ranging from 0.00 – 0.15, 0.00 – 0.30 and 0.00 – 0.17, respectively. Additional results, such as the selected model, optimal hyperparameter settings, and evaluation metrics for each HRV measure, can be found in Table A.4 of Appendix A.7.

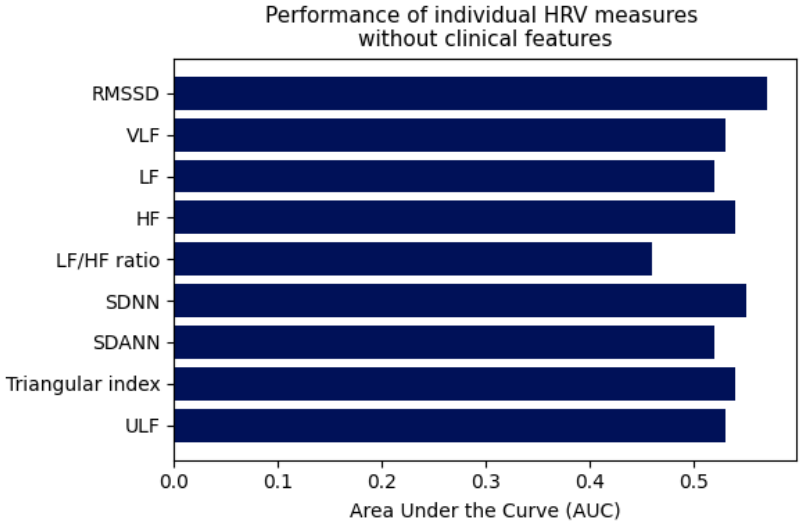



**Figure 1.6: Performance of the individual HRV measures on the prediction of ICU mortality expressed as AUC.**

The inclusion of clinical features improved performance. The performance of the clinical features combined with individual HRV measures yielded AUCs ranging between 0.70 and 0.73, with the power in the HF band achieving the best performance (Figure 1.7). This, however, was a marginal improvement over when only clinical features were used, which resulted in an AUC of 0.72. The additional results of these combinations are presented in Appendix A.7 (Table A.5).

Next, we explored whether combinations of HRV measures could improve performance, as the correlation analysis showed that they capture different aspects of the HRV. We tested all pairwise combinations of HRV measures together with the clinical features to see how this would affect the performance. AUCs ranged between 0.68 and 0.74 (Figure 1.7). The best performing model combined the clinical features with the power in the HF band measured over five minutes within the first hour of admission and the SDANN measured over the first eight hours of admission. Again, in terms of AUC, the performance gain is only 0.02 compared to using only clinical features without HRV. Finally, we tested three additional combinations: all short-term HRV measures, all long-term HRV measures and all HRV measures combined with the clinical features. For these combinations, we did not include the highly correlated HRV measures as discussed in Section 2.3.2. The results are shown in Figure 1.8. The combination of the short-term HRV measures performed best, achieving an AUC of 0.76. The HRV measures included in this short-term set were the power in the VLF band, the power in the HF band and the LF/HF ratio. We constructed the ROC curves for this combination to show the performance across the ten folds (Figure 1.9). The results of the other combinations are shown in Table A.5 in Appendix A.7. The model that was selected for this feature set was an

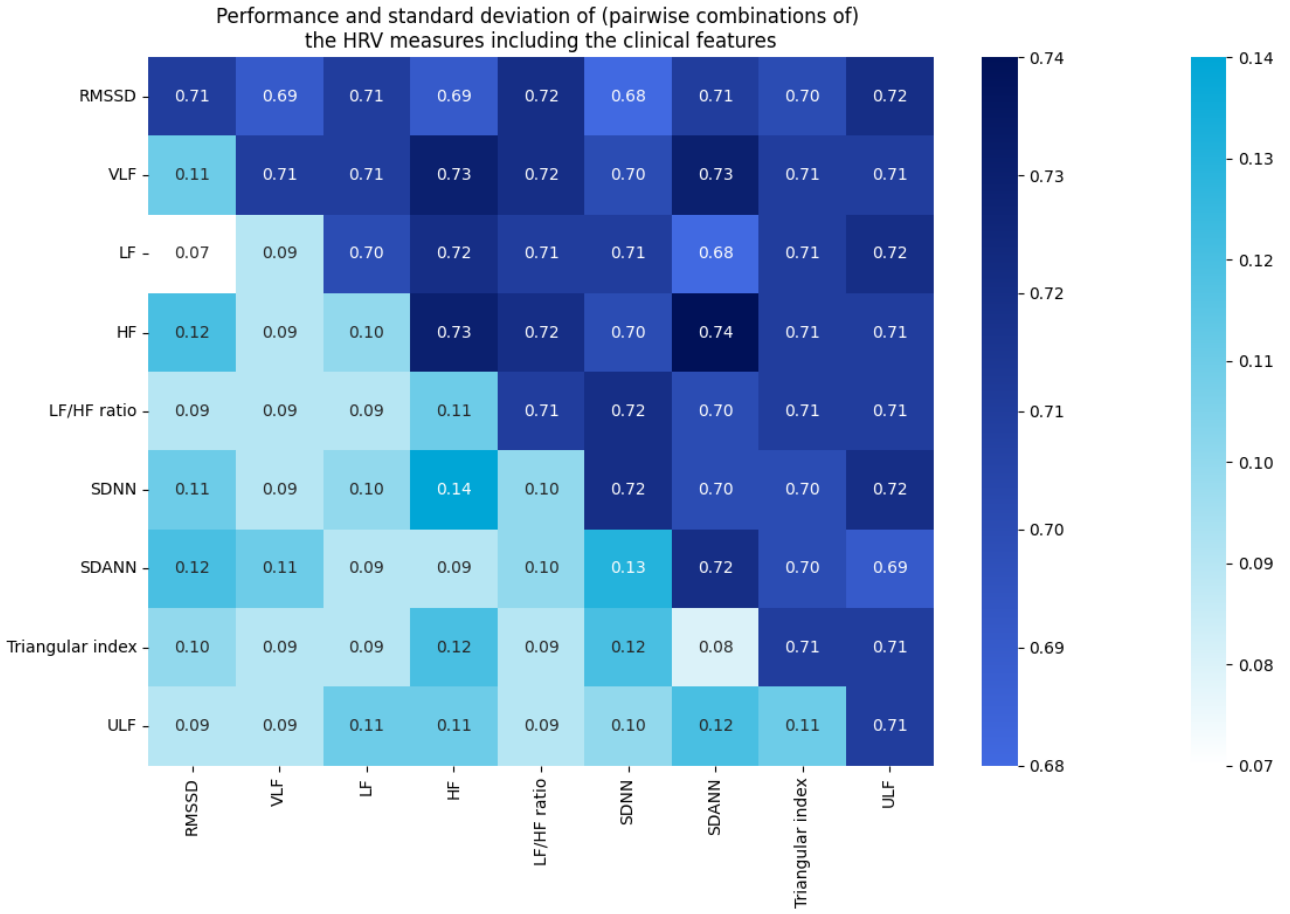



**Figure 1.7: The performance of (pairwise combinations of) HRV measures combined with clinical features expressed as AUC.** The diagonal shows the performance of each individual feature. The right side shows the best estimated performance of the pairwise combinations, while the left side shows the standard deviation of each combination, as the performance is determined by averaging over ten folds.

XGBoost model. Feature importance scores and SHAP values were calculated for this model (Figures 1.10 and 1.11). Both the feature importance scores and the SHAP values indicated that the APACHE IV score contributed the most to the prediction. Regarding the HRV measures, the power in the VLF band and the power in the HF band contributed the most, with the SHAP values indicating that they had the second and third highest impact on model performance right after the APACHE IV score.

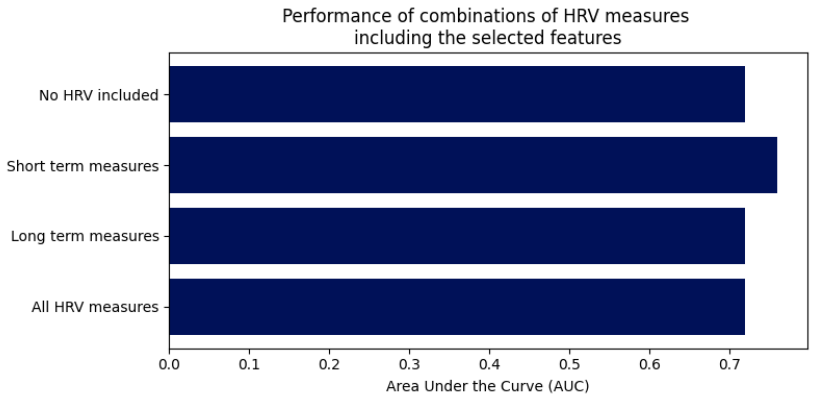



**Figure 1.8: The performance of combinations of HRV measures (>2) including clinical features.**

The other evaluation metrics—sensitivity, precision, and F1 score—showed poor outcomes for the

feature sets that combined individual HRV measures or various combinations of HRV measures with clinical features, although they did increase compared to the outcomes using only individual HRV measures. Sensitivity ranged from 0.43 to 0.45, precision ranged from 0.13 to 0.45, and the F1 score ranged from 0.33 to 0.40 (Table A.5 in Appendix A.7).

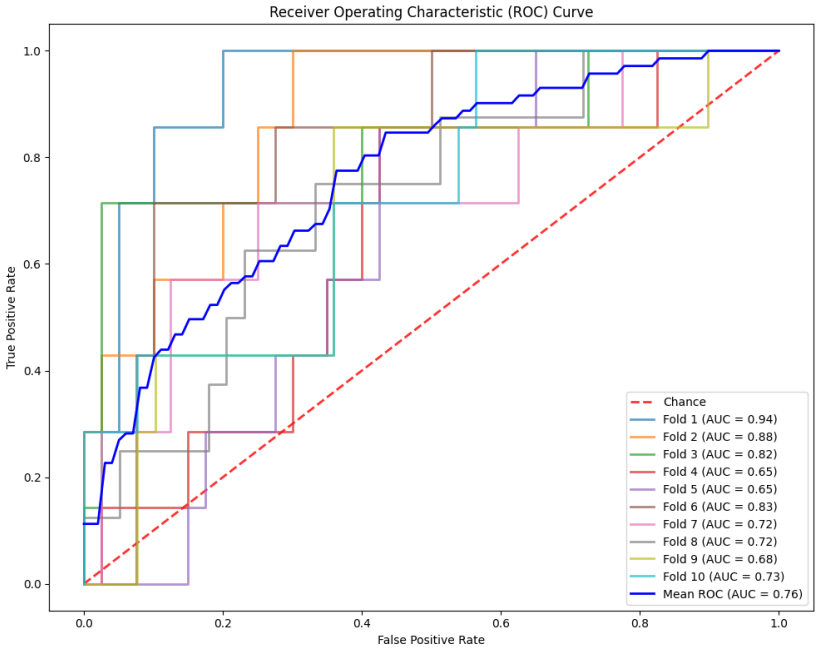



**Figure 1.9: ROC curves showing the AUC for each of the 10 folds and their mean for an XGBoost model combining all clinical features with the short-term HRV measures.**

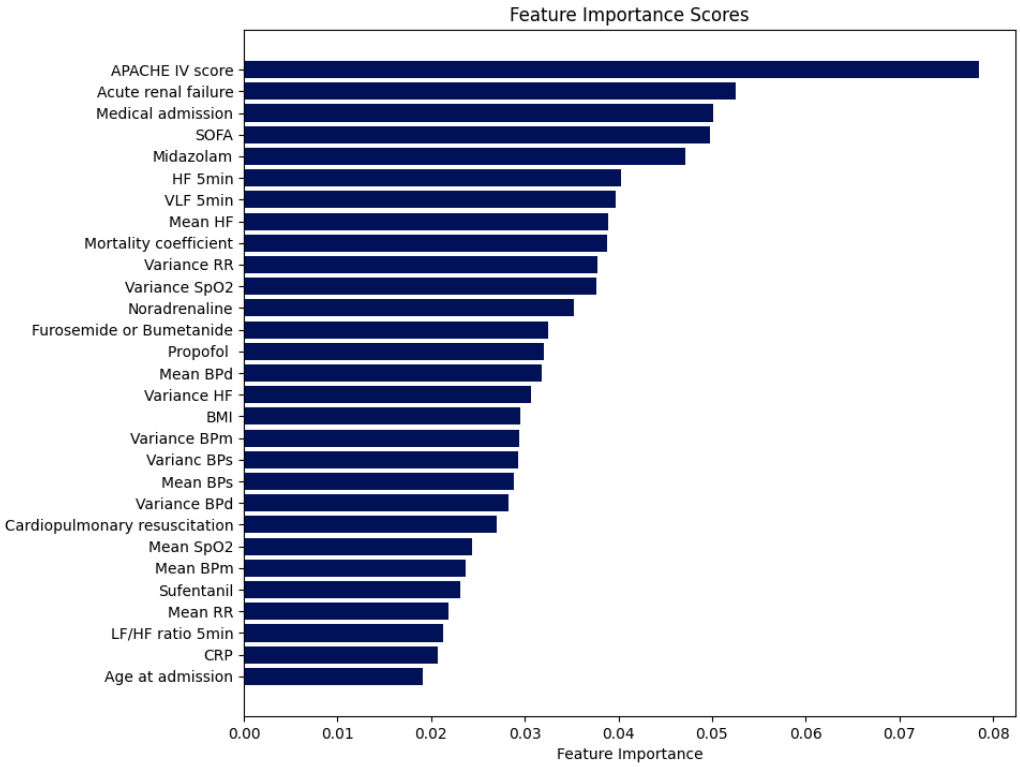



**Figure 1.10: Feature importance plot of the XGBoost model combining all clinical features with the short-term HRV measures.**


**Figure 1.11: SHAP values of the XGBoost model combining all clinical features with the short-term HRV measures.**

### 1.4.4 Predictive models for intensive care unit length of stay

For the prediction of ICU length of stay, we explored three different regression models: a Lasso regression model, a SVR model and an XGBoost regressor. We used the same approach for model selection and performance estimation as we did for predicting ICU mortality, i.e., nested cross-validation (Figure 1.4).

We decided to perform the LOS prediction solely on ICU survivors because including patients with a relatively short LOS who died might incorrectly bias the model towards considering them as 'healthier'. To reduce this bias, we excluded the 71 ICU non-survivors from this analysis. For non-survivors, we created two plots to illustrate the timing of deaths in the ICU, showing that a large portion of deaths occur within the first ten days. These plots are shown in Figures A.3 and A.4 in Appendix A.8.

Predicting ICU LOS using only single HRV measures yielded poor results, with MAEs varying from 5.15 to 5.20. A SVR using the power in the ULF band achieved the lowest MAE, making it the best-performing model. Additional results, including the best regressor for each feature set and the out-

comes of other evaluation metrics, are presented in Table A.6 in Appendix A.9.

When clinical features were added, the performance in predicting ICU LOS using individual HRV measures improved only very slightly, yielding MAEs ranging from 5.08 to 5.10 (Figure 1.12). We found that a support vector regressor was the best model for each combination of an individual HRV measure with clinical features. Using clinical features alone, the best predictive model yielded an MAE of 5.09, indicating that HRV features did not substantially improve performance.

Next, we examined whether pairwise combinations of HRV could improve the performance of the prediction models. The MAEs ranged between 5.07 and 5.10, so no real performance improvement was found (Figure 1.12).

Finally, we explored the use of the combination of clinical features with all short-term HRV measures, all long-term HRV measures and all HRV measures, yielding MAEs of 5.07, 5.10 and 5.08, respectively. All the outcomes regarding the prediction of ICU LOS, including the best regressor for each feature set and the outcomes of other evaluation metrics, are listed in Table A.7 in Appendix A.9.

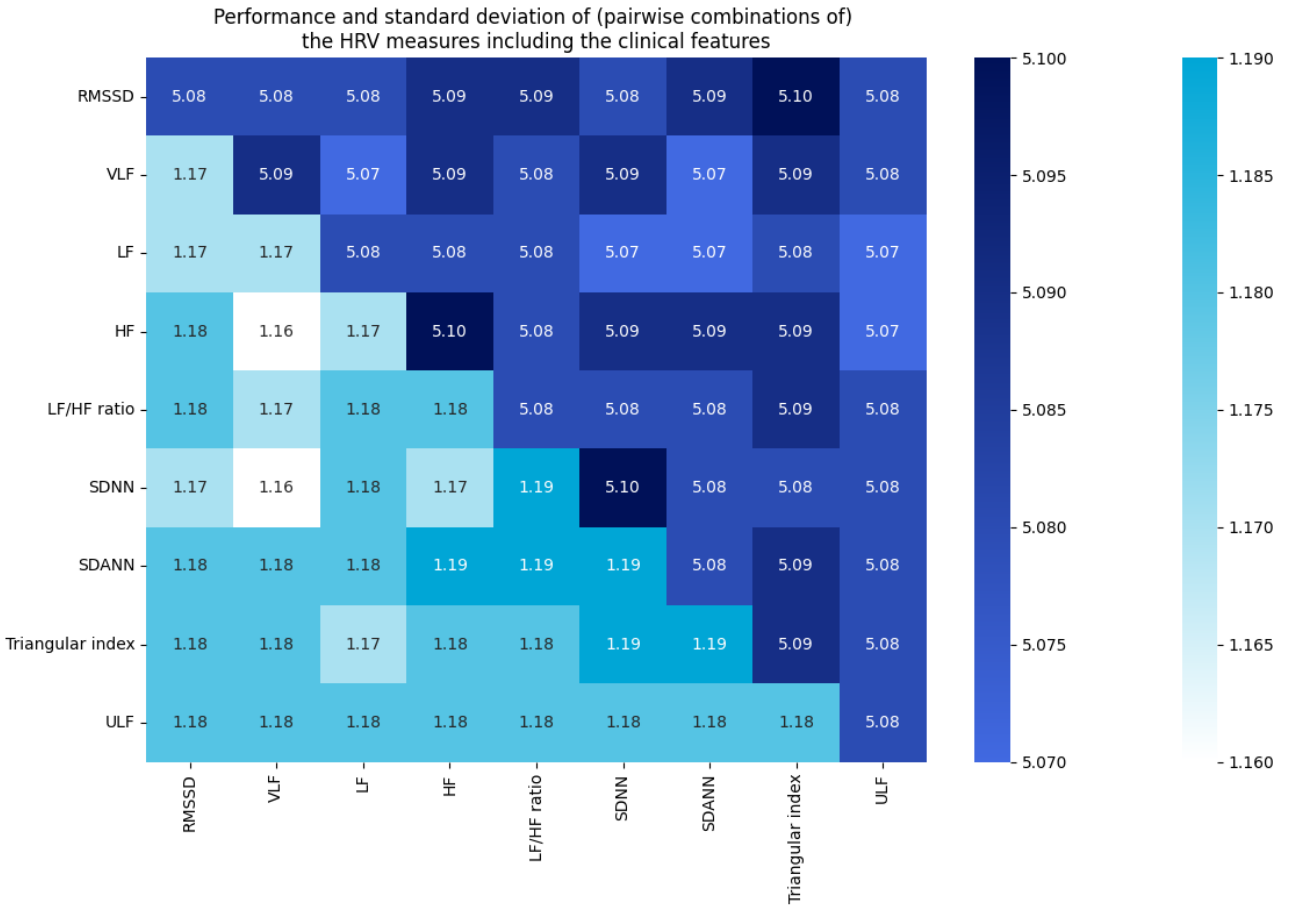



**Figure 1.12: The performance of (pairwise combinations of) HRV measures combined with clinical features expressed as MAE.** The diagonal shows the performance of each individual feature. The right side shows the best estimated performance of the pairwise combinations, while the left side shows the standard deviation of each combination, as the performance is determined by averaging over ten folds.

As we observed only small differences between using HRV measures alone and adding clinical features, and the baseline using clinical features alone improved only marginally when HRV measures were added, we decided to establish another baseline. This would allow us to assess whether the HRV measures and/or clinical features had any meaningful effect on model performance. To do this,

we used a common baseline in regression problems, often referred to as a dummy regressor: the mean of the target variable, which in our case was LOS (46). This baseline resulted in an MAE of 6.21, demonstrating that our model outperformed the dummy regressor.

The RMSE and $R^2$ values also showed poor performance. Without clinical features, using only HRV measures, RMSE ranged from 9.84 to 15.32 and $R^2$ ranged from -0.16 to -0.04. When clinical features were included, RMSE ranged from 9.53 to 9.95 and $R^2$ ranged from -0.04 to 0.03 (Table A.6 and Table A.7 in Appendix A.9).

As there was no clear set of features that achieved the best performance, we decided to determine the feature importance and SHAP values for the best performing model using all clinical features and all uncorrelated HRV measures. This was a support vector regressor. As a support vector regressor does not support native feature importance scores, we used permutation feature importance. This is a technique that can be applied to any model to calculate the relative importance of features, and is particularly useful for models that do not provide native feature importance scores (47). The feature importance score plot and the SHAP value plot are shown in Figures A.5 and A.6 of Appendix A.10.

## 1.5 Discussion

We studied whether heart rate variability was predictive of (short-term) clinical outcomes in ICU patients. We employed a machine learning approach to assess the predictive value of nine HRV measures on ICU mortality and ICU LOS. When HRV measures alone were used to predict mortality, poor results were found, with AUCs ranging between 0.46 and 0.57. Models that combined clinical features with individual HRV measures or combinations of HRV measures improved performance and were considered fair (AUC 0.70) for the prediction of ICU mortality (48). The best performing model was an XGBoost classifier that combined three short term HRV measures (power in the VLF and HF band and LF/HF ratio) with the clinical features and achieved an AUC of 0.76. With a mean absolute error of approximately 5 days, the regression models used to predict ICU LOS performed poorly. Little difference was observed between the various combinations of features, suggesting that HRV did not contribute significantly to the models.

### 1.5.1 Interpretation of the results

#### 1.5.1.1 Heart rate variability measures

We observed a strong correlation between RMSSD and the HF band which has been previously reported in the literature (Figure 1.5) (41, 42). This correlation can be attributed to their shared reflection of parasympathetic activity. The observed correlation between the SDNN and the SDANN has also been frequently reported in the literature and can be explained by the fact that the measures are based on almost the same mathematical formula (42, 43). SDNN computes the standard deviation of all normal RR intervals over a given period, whereas SDANN averages normal RR intervals in 5-minute segments before calculating their standard deviation. As a result, the SDANN is less susceptible to editing errors than the SDNN because the averaging of RR intervals minimizes the effects of unedited artefacts, missed beats and ectopic complexity (42). The correlation between the power in the low frequency band and the power in the high frequency band has also been previously reported, but requires a bit more explanation (41). In Table 1.1, we reported that LF power is

influenced by both the sympathetic and parasympathetic nervous systems. However, there is a lot of scepticism about the representation of sympathetic activity in the LF band. An important observation was that vagal blockade significantly reduced LF power, whereas sympathetic blockade showed minimal effect (49). Miki et. al. indicated that across a broad frequency spectrum including LF and HF bands, spectral power indices were predominantly influenced by the parasympathetic nervous system (50). Therefore, the correlation between the LF and HF power was understandable because both reflect the dominance of parasympathetic activity.

#### 1.5.1.2 Predictive models for intensive care unit mortality

The models that used only the individual HRV measures did not perform well, with AUCs ranging from 0.46 to 0.57 (Figure 1.6), suggesting that HRV alone has little predictive value for ICU mortality. The AUC increased when combined with carefully selected demographic and clinical features. Compared to a baseline model, using only clinical features (AUC of 0.72), certain combinations of HRV measures did lead to increased performance. The feature set with the best performance included the clinical features and all short-term HRV measures, achieving an AUC of 0.76. Especially the power in the HF band and the power in the VLF band appeared to contribute to this performance (Figures 1.10 and 1.11). Both measures were calculated based on 5 minutes in the first hour of admission and were higher in patients who died during ICU admission (Table 1.4). High power in the HF band indicates increased parasympathetic activity. The VLF, especially when calculated over 5 minutes, remains challenging to explain physiologically. Apart from the thermoregulatory and hormonal mechanisms reported in Table 1.1, the literature offered several other interpretations. For example, VLF has been reported to be associated with increased chronic inflammation and it has been suggested that the VLF rhythm is intrinsically generated by the heart and modulated by efferent sympathetic activity (9, 13, 51).

In addition, the combination of HF power and SDANN with the clinical features increased the AUC from 0.72 to 0.74 (Figure 1.7). This was an interesting observation as it combined a short-term frequency domain measure with a long-term time domain measure. This suggested that a short-term reflection of parasympathetic activity along with a long-term reflection of the ANS - including circadian rhythm, temperature, metabolic and hormonal systems – outperformed the baseline model, which included only the clinical features and no HRV, in predicting ICU mortality.

We used the AUC as the primary evaluation metric for predicting ICU mortality. However, there appeared to be no consensus in the literature on the interpretation of this value. For instance, de Hond et al. found that there is a wide variety in the AUC labelling system in research papers (52). When comparing our highest AUC of 0.76 with the literature, it was described as ranging from moderate, fair, or acceptable to good to very good, high or excellent. In their paper, they highlighted that achieving high discriminatory ability alone is not sufficient to claim a positive potential impact of using a prediction model in clinical practice. The primary role of AUC values is often to compare the discriminatory abilities of different models. Our study focused on assessing the predictive value of HRV, and in the future, we may consider further developing the model that achieved the highest AUC.

Nevertheless, we assessed our models using other evaluation metrics, i.e., sensitivity, precision and the F1-score. We included sensitivity, which measures the model's ability to correctly classify a person as deceased during ICU admission, because it is a measure commonly used in the medical

world and therefore intuitive for physicians to interpret. Precision is the proportion of true positive predictions out of all positive predictions made by the model (53). The F1-score combines precision and sensitivity to provide a single measure of a model's accuracy, especially useful for imbalanced datasets where accuracy is often misinterpreted because it largely disregards the performance of the minority class (54). A low sensitivity indicates inadequate identification of actual ICU deaths, while a low F1 score reflects overall poor performance of the binary classification model due to low sensitivity and precision. Despite having a reasonable AUC, the low sensitivity, precision and F1-score of our models suggested challenges in accurately identifying actual ICU deaths, potentially influenced by dataset imbalance.

#### 1.5.1.3 Predictive models for intensive care unit length of stay

The models for predicting length of stay performed poorly. All regression models showed an MAE of approximately 5 days, regardless of whether they used individual HRV measures alone or (combinations of) HRV features with clinical features. To assess whether the HRV and clinical data affected the predictions, we evaluated a dummy regressor, which resulted in an MAE of 6.21. This indicated that clinical features and HRV measures had some impact, albeit minimal and not sufficient for implementation.

Like the MAE, the RMSE indicates the predicted discrepancy in days. The $R^2$ value indicates the percentage of variation in the dependent variable that is explained by the regression model (55). A value of 0 means that the model explains none of the variation in the dependent variable, indicating that the model does not perform better than using the mean of the dependent variable. A value of 1 means that the model explains all the variation in the dependent variable. All the $R^2$ values we observed were around 0. The best achieved RMSE of 9.5 days and the $R^2$ values around zero, confirmed that the individual HRV measures or HRV measures combined with clinical features were not a good predictor of ICU length of stay.

### 1.5.2 Comparison with the literature

It is often hypothesized in the literature that HRV is lower in critically ill patients due to autonomic dysfunction, as a low HRV indicates reduced adaptability of the heart and autonomic nervous system to changing physiological demands, meaning that the body's ability to respond to stress is impaired (56). However, we also found evidence that some HRV measures may be higher in critically ill patients. For instance, Chen et al. demonstrated a significant increase in normalized HF strength (HF power relative to the total power) in patients at risk of death, establishing it as the most reliable predictor of mortality in ARDS patients (33, 57). They concluded that in critically ill patients following thoracic surgery, increased vagal modulation may be an indicator of a poor prognosis. We found a similar result in our study, where power in the HF band was significantly increased in ICU non-survivors (Table 1.4). In addition, of all the models that predicted mortality with the clinical characteristics and a single HRV measure, the model that included power in the HF band performed best with an AUC of 0.73 (Figure 1.7). Furthermore, power in the HF band was part of the best performing model and had the highest feature importance and SHAP values (Figures 1.10 and 1.11), indicating its substantial contribution to the prediction.

In addition to the contribution of power in the HF band, we found that power in the VLF band also substantially improved the performance of the best model (Figures 1.10 and 1.11). This is con-

sistent with the existing literature, which reports that VLF power is more strongly associated with all-cause mortality compared to power in the LF or HF bands (9, 13).

The SDANN measure also showed a significant difference between ICU survivors and non-survivors (Table 1.4). Few studies investigating the predictive power of HRV in a critically ill population have included SDANN as an HRV measure, as SDNN is often preferred and the two are strongly correlated. Nevertheless, there have been studies investigating the effect of SDANN in other populations. Swearingen et al. published an abstract showing that lower SDANN is a strong predictor of heart failure hospitalization within six months of cardiac resynchronization therapy defibrillator implantation (58). Di Franco et al. also showed that the SDANN was significantly lower in patients with systemic sclerosis than in healthy controls (59). These results, in which the impaired group had a significantly lower SDANN, are consistent with those of our study.

To the best of our knowledge, there have been no previous studies predicting ICU LOS using HRV measures. However, there are many studies that have attempted to predict ICU LOS using other parameters. For example, Hempel et al. presented predictive models for LOS in ICU patients using demographic, administrative and early clinical data from the MIMIC-IV database (60). They concluded that the models performed poorly when considering a maximum LOS of 21 days. They suggested that this may be because the models only included data from the first day of ICU admission, which limits their ability to accurately predict longer LOS durations. This may also explain the poor regression results in our study, as we only used data from the first 24 hours of admission as well. In future research, it would be interesting to investigate HRV trends throughout the ICU stay.

### 1.5.3 Clinical implications

During this prognostic study, we aimed to determine whether HRV possesses the ability to predict short-term outcomes in the ICU. Patients in the ICU are continuously monitored, and clinicians make ongoing decisions based on the measured values. HRV would be a valuable addition to existing parameters, as it provides us with a measure of the balance between the two arms of the autonomic nervous system. In the ICU, it is particularly important to identify patients at high risk of mortality, and less important to classify low-risk patients as such. Therefore, having a model with high sensitivity is important. The models presented in this study are not yet ready for clinical implementation. However, we demonstrated that including HRV measures slightly improved the performance of the model that predicted mortality using only clinical features. The results gave us confidence that in the future, with further research, models using HRV can be developed that can identify high-risk patients in the ICU.

### 1.5.4 Strengths and limitations

One of the strengths of our study was that we spent a lot of time developing an accurate ECG processing algorithm. This algorithm used several evidence-based methods (Figure 1.3). A limitation was that there was insufficient time to perform an accuracy analysis, such as manually marking R-peaks and comparing them with the R-peaks correctly detected by the algorithm. However, the development of a robust artefact detection algorithm was not the focus of this study, but merely a necessary tool, and therefore outside the scope of this study.

Another strength was that we assessed the predictive power of multiple feature sets to understand both the individual contribution of the HRV measures as well as their combinations Another strength

was that we analysed multiple models to understand both the individual contribution of the HRV measures as well as their combinations as effectively as possible. Although the AUCs showed little difference, even a modest increase in the AUC can provide meaningful insights into the predictive ability of additional parameters in risk assessment (61, 62).

The size of our study dataset can be both interpreted as a strength and a limitation. Compared to other HRV studies, our dataset was relatively large (Appendix A.3). However, it was considered small for a machine learning study. One reason for the limited dataset size was the availability of the high-frequency data from the Data Warehouse, which was only accessible for one year before being overwritten due to capacity reasons. In addition, patients who stayed in the ICU for less than 48 hours were excluded, further reducing the number of patients. A rule of thumb in machine learning is that "a dumb algorithm with lots and lots of data beats a clever one with modest amounts of it" (63). We therefore expect that increasing the amount of data would improve the performance.

A limitation was that we calculated the long-term HRV measures over 8 hours instead of the recommended 24 hours. Unfortunately, this was unavoidable due to computational issues caused by the large number of data points from the 500 Hz sampled ECG. Given that the 24-hour data was available, a solution would be to run the HRV calculations on a computer with greater computational power.

Another limitation was that some ECG data might have been missing due to (re)surgery or other reasons for disconnection from the monitor. It was also possible that some patients might have had more five-minute segments excluded due to poor quality compared to others, which could result in HRV being calculated over varying numbers of RR intervals across patients.

A limitation of our regression model is that we chose to exclude ICU non-survivors from the LOS prediction to avoid potential bias. We made this decision because a significant proportion of ICU non-survivors died within the first 10 days of admission (Appendix A.8). Including them could have biased the model by interpreting their shorter LOS as an indication that they were 'healthier', potentially distorting the predictions for other patients. An alternative approach would have been to use a survival analysis model, which explicitly takes into account censored data, such as patients who died before hospital discharge (64). By using survival analysis, we could have modelled the time to event (ICU discharge or death) more appropriately, potentially leading to a more accurate prediction of LOS across all patient groups

### 1.5.5 Recommendations for future research

Our machine learning approach, which used nested cross-validation to prevent overfitting and ensure robust performance estimation, was a suitable method for predictive research. We therefore recommend its continued use. However, we recommend that a future study be conducted using a larger dataset. We expect that this will improve the performance of the model by providing more data for learning. In addition, we suggest including more HRV measures, including non-linear ones, to gain a more comprehensive understanding of HRV and its reflection of the autonomic nervous system. Finally, the use of a computer with greater processing power would facilitate the calculation of HRV over 24 hours, allowing a better comparison with the literature, which often discusses 24-hour rather than 8-hour measures.

## 1.6 Conclusion

In combination with other clinical features, HRV contained substantial predictive information for ICU mortality. Promising HRV measures for predicting ICU mortality were power in the HF band, power in the VLF band and SDANN, consistent with previous findings. Models using clinical features with or without HRV measures were not able to accurately predict ICU length of stay. Further and more extensive research into the use of HRV to predict both ICU mortality and ICU length of stay is essential before implementation in clinical practice can be considered.

It is worth noting that although it is commonly assumed that HRV is reduced in critically ill patients, this is not always the case. HRV should not be perceived as simply higher or lower, but rather as a deviation from typical patterns.

# 2

# Exploring feasibility of eHealth monitoring in Intensive Care Unit survivors


## Abstract

**Background:** Advances in technology have improved ICU survival rates, but half of these survivors suffer from Post Intensive Care Syndrome, which negatively impacts their quality of life and increases their healthcare needs. A pilot study demonstrated the feasibility of home monitoring for ICU survivors using smart technology, but with only fifteen participants, a larger clinical trial is needed to assess feasibility on a larger scale.
**Objective:** The aim of this study was to prepare for the implementation of the ICU Recover Box Study 2.0.
**Methods:** We evaluated the results and challenges of pilot study 1.0 and used the lessons learned to create a new study design. We arranged all the necessities for the Medical Ethical Review Commission (METC) application and thought out the entire study logistics. This included conducting thorough research into the study trajectories and creating standard operating procedures while maintaining contact with external parties and other departments within the hospital that were involved.
**Results:** Although preparation is largely operational and difficult to quantify, we prepared thoroughly for the start of the study and gained valuable insights into clinical research. Three key takeaways are that research is an iterative process with ongoing learning, the study objective must be carefully considered with ethical concerns in mind, and the growing importance of data requires careful planning for its security, processing and storage.
**Conclusion:** The METC application has been submitted. Once approved, the study will be ready to proceed on a well-prepared basis.

## 2.1 Introduction

Due to advances in technology and practice, an increasing number of critically ill patients survive the ICU and are discharged to their homes (3–5). Despite improved survival rates, the ICU remains associated with mortality and morbidity. Patients may continue to experience the consequences of an ICU admission even after discharge. These consequences are enclosed in the term Post Intensive Care Syndrome (PICS) which is defined as "worsening impairments in physical, cognitive, or mental health status arising after critical illness and persisting beyond acute care hospitalisation" (65). One in two patients who survive a critical illness is affected by PICS (66). Despite the heterogeneity of the syndrome, patients suffering from PICS have an increased likelihood of reduced quality of life, hospital readmission or even death (66–68).

Several studies have investigated the outcomes of ICU survivors after one year. For instance, Gonçalves-Pereira et al. reported a 6.3-fold increase in the risk of death in the first year after ICU discharge (6). In a systematic review, McPeake et al. found a pooled estimate of readmission after critical illness at 12 months to be 53% (69). They identified several patient-specific risk factors for readmission, including comorbidities, frailty, and the specifics of the initial hospitalization. Prescott et al. confirmed the high readmission rate, describing a readmission rate of 43% at 90 days among survivors of severe sepsis (70). Prescott et al. also concluded that up to 42% of the readmitted patients were admitted with diagnoses that could have been prevented (70). In addition to the inconvenience for the patient, there are high healthcare costs associated with persistent morbidity and healthcare utilization after discharge from the ICU.

These numbers demonstrate that, in addition to the need for accurate monitoring to identify high-risk patients during ICU admission, it is also crucial to identify patients at increased risk for complications after ICU discharge. A post-ICU outpatient clinic is not yet part of standard care in the Netherlands, although some hospitals offer this care (71). Concerns remain about the cost-effectiveness of these clinics (72). Remote monitoring could be an ideal compromise, allowing doctors to spend less time while still monitoring and managing patients effectively. They can intervene promptly, for example by referring patients to their general practitioner if problems arise. This could potentially help reduce healthcare utilization numbers, reducing both the pressure on the healthcare system and the economic burden.

## 2.2 Background

### 2.2.1 Systematic literature review

We conducted a systematic review of the healthcare utilization of ICU survivors in the year after discharge. This review is included in Appendix B.1. The main finding is that health care use is substantial among ICU-survivors in the year after discharge. In over 80% of the studies, more than a third of the ICU-survivors had to be readmitted to the hospital. Readmissions to the ICU and visits to the emergency department or other health care professionals were also common. The costs associated with health care use in the year after ICU discharge were high.

### 2.2.2 Home monitoring and eHealth

In 2022, Viderman et al. published a systematic review on remote monitoring of chronically critically ill patients after hospital discharge (73). They explain that the population of chronically critically ill patients continues to grow due to advances in intensive care medicine. Wearable devices offer significant support to medical staff and caregivers by monitoring vital signs like blood pressure, heart rate, respiratory rate, blood oxygen saturation, metabolism, and central nervous system function. Despite extensive research into telemonitoring, many uncertainties remain, particularly in terms of device performance, safety, clinical and economic outcomes, and acceptance by patients and healthcare professionals. Therefore, further research is necessary to address these uncertainties and advance our understanding of remote monitoring technologies.

In 2020, Treskes et al. published the results of a randomized clinical trial investigating whether smart technology in clinical practice could improve blood pressure regulation and its feasibility in the year after myocardial infarction (74). The main finding of this study was that monitoring patients with smart technology and changing two out of four outpatient visits to electronic visits resulted in similar percentages of patients with regulated blood pressure compared with standard care (i.e., no use of smart technology and four outpatient visits in the first year after myocardial infarction). In addition, monitoring patients with smart technology was feasible and deemed acceptable to patients. In 2022, they published another study on the same topic to determine the cost-effectiveness of such an intervention (75). They found that using eHealth in the outpatient setting for myocardial infarction patients is likely to be cost-effective compared with regular follow-up. Although the population of our study is different, the results of these studies show promising results for the general use of eHealth in patient follow-up.

### 2.2.3 Pilot study 1.0

A team of researchers from the ICU of the LUMC conducted a pilot study, in which (vital) parameters of fifteen ICU survivors were monitored after discharge from the hospital (76). The aim of the pilot was to determine the feasibility of monitoring ICU survivors at home using devices that measure heart rate, blood pressure, weight, temperature, oxygen saturation and activity. Patients were also given questionnaires about their quality of life (QoL) and healthcare use. The pilot study showed that home monitoring of ICU survivors was feasible, as more than 80% of the enrolled patients provided their data throughout the study period, and the data could be collected from the devices and stored in a secure manner.

## 2.3 Methods

The aim of this part of the thesis was to prepare for the implementation of a follow-up study, i.e., the ICU Recover Box 2.0. In this section, we will briefly discuss the methods of this clinical study and then discuss the approach to preparing for this study.

### 2.3.1 The ICU Recover Box 2.0

#### 2.3.1.1 Objective

The primary objective of the ICU Recover Box 2.0 study was to assess the feasibility of having ICU survivors use smart technology in the year following hospital discharge. Smart technology refers to

the Corsano CardioWatch 287-2 and the accompanying blood pressure band that make up the ICU Recover Box. Feasibility was defined as:

- 50 ICU survivors (who gave informed consent) were discharged from the ICU with the ICU Recover Box.
- 80% of the included patients were able to use the devices within their intended use.
- We were able to acquire data from the Corsano CardioWatch 287-2.
- We were able to check the data regularly on reliability and validity.
- We were able to act on missing data.
- We were able to store the acquired data in a secure manner according to the Medical Device Regulation (MDR) and LUMC regulations.
- We were able to analyse the acquired data.
- 80% of the included participants contributed to post-ICU data for one year.

Secondary objectives of the study were:

- To systematically collect all data on telephone calls or emails from participants regarding the use of the Corsano CardioWatch 287-2 and to create a comprehensive overview of the frequency and types of issues encountered by users.
- To conduct standardized oral interviews (via telephone) with participants to gather feedback, allowing them to suggest adjustments and improvements.
- To use the Plan-Do-Check-Act cycle to incorporate lessons learned and feedback, thereby improving all aspects of the ICU-Recover Box—from CardioWatch usage to data collection and analysis—for future studies or clinical applications.

#### 2.3.1.2 Study design

The ICU Recover Box 2.0 is a longitudinal, prospective, single-centre pilot study that has not yet started. We can divide the study into three phases: the preparation phase, the study phase, and the post-study phase. The preparation phase is needed to ensure that everything is in place for the study to start, from obtaining study approval from the Medical Ethical Review Commission (METC) to creating standard operating procedures for conducting the study. During the study phase, patients will be enrolled in the study and followed at home for one year after enrolment. Questionnaires will be conducted according to the schematic overview in Figure 2.1. The post-study phase consists of data processing and analysis. Table 2.1 provides an overview of the activities for each phase.

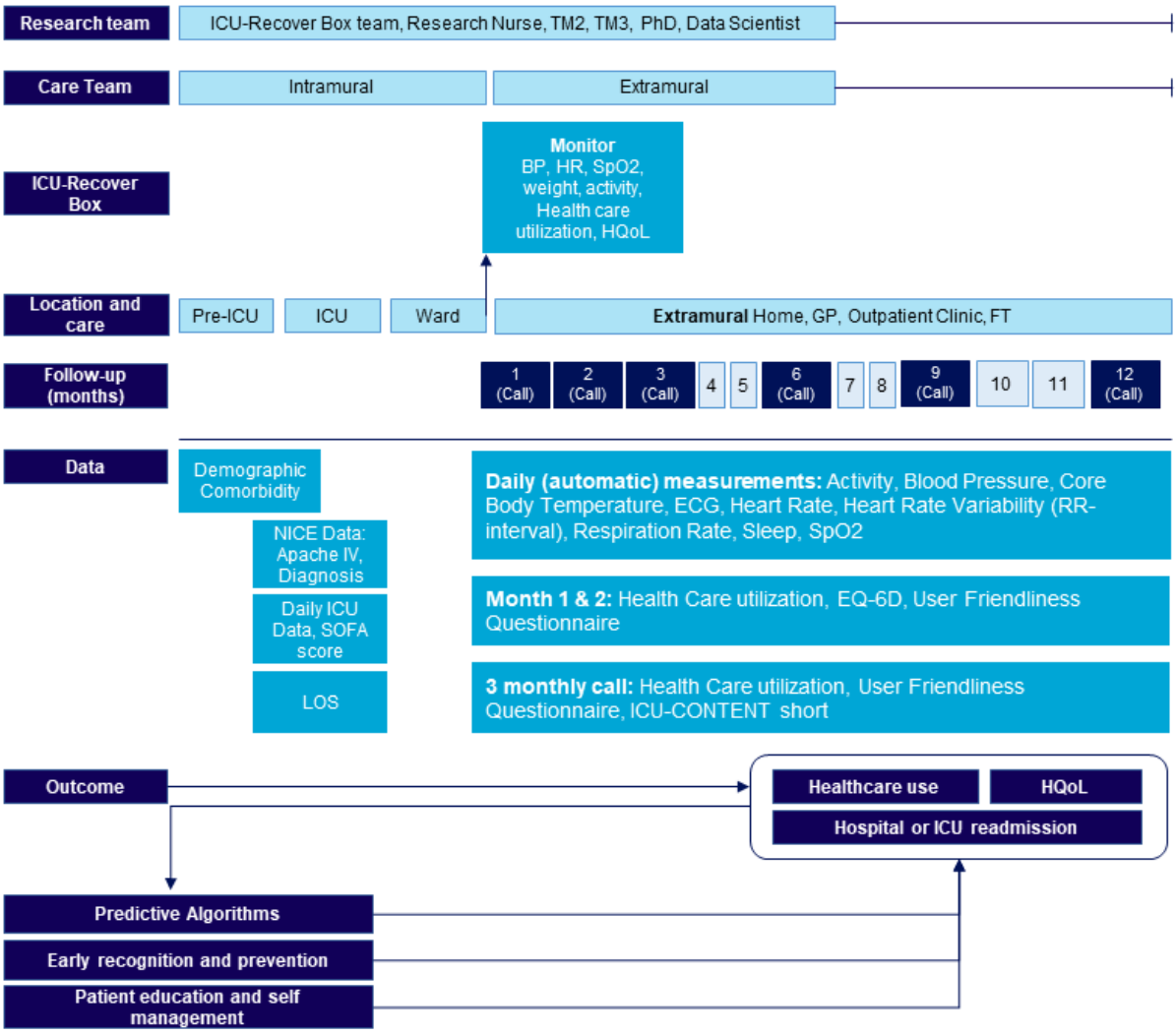



**Figure 2.1: A schematic overview of the ICU Recover Box 2.0 study.**

#### 2.3.1.3 Subjects

All patients admitted to the ICU of the LUMC from the start of the study phase onward will be reviewed. Inclusion criteria are:

- Patient has been admitted to the ICU for 24 hours.
- Patient has received mechanical ventilation during the ICU admission.
- Patient masters the Dutch or English language.
- Patient is capable of using smart technology at home.
- Patient is discharged from a nursing ward of the LUMC to home or to a rehab facility.

Participants that meet any of the following exclusion criteria will be excluded from participation in the study:

- Patient is <18 years old.
- Patient is pregnant.
- Patient breastfeeds during the course of the study.
- Patient is discharged for palliative care.
- Patient is considered an incapacitated adult.
- Patient is unwilling to sign the informed consent form.
- Patient is discharged or transferred to another hospital.

**Table 2.1: An overview of the activities per study phase.** ICU: Intensive Care Unit.

| Preparation | Study | Post-study |
|---|---|---|
| Evaluate the results and challenges of pilot study 1.0. | Monitor all ICU patients for study eligibility. | Process and analyse the study data. |
| Create a new study design, including reassessment of the monitor device(s). | Get informed consent and include patients in study. | Create an overview of the results: feasibility and secondary objectives. |
| Write the study protocol and ensure that all the necessary documents are ready for the METC application. | Monitor patient data through research protocol. | Put together a list of recommendations for future research. |
| Conduct thorough research into the study trajectory for both the patient and the researchers. | Contact patients in case of abnormalities, such as when measurements fail to come through or there is a deteriorating trend. | Put together a list of recommendations for implementation of the ICU-Recover Box in standard care. |
| Create standard operating procedure for conducting the study. | Conduct questionnaires by phone. | Publish results in a scientific journal. |
| Make decisions about the frequency of data retrieval and manual measurement of patients. | Document all instances of patient contact, both scheduled and unscheduled. | |
| Adapt the questionnaires to the revised objectives of pilot study 2.0. | | |
| Maintain contact with external parties and other departments within the hospital. | | |
| Prepare informed consent procedure and forms. | | |
| Prepare the contents of the boxes, including all devices and information for the patient. | | |
| Prepare the essentials for online patient monitoring, including the research portal and communication tools such as email and a phone. | | |

The aim is to include 50 patients. After discharge from the ICU, patients will be admitted to the nursing ward where they will be contacted before being discharged to home or another (rehabilitation) facility. All potential participants will be given a verbal overview of the study and a detailed information sheet explaining the purpose and activities of the study. The researcher must ensure that patients have the opportunity to ask questions. Written informed consent must be obtained from each patient. This study is subject to WMO and therefore requires METC approval.

#### 2.3.1.4 Home monitoring device

The largest change that was implemented in the ICU Recover Box 2.0 compared to pilot study 1.0 is the use of a different smart technology device. In pilot 1.0, the devices in the ICU Recovery Box were the Withings ScanWatch, Withings Body and Withings BPM Connect. There were two issues with these devices. The first was that the Withings ScanWatch and Withings Body were not CE marked. This resulted in the METC imposing additional precautions on the researchers, such as the researchers having to be available 24 hours a day, seven days a week for (medical) questions from the participants. Secondly, the ScanWatch was found to be difficult to use in patients with

neuromuscular impairment, which is a common condition in ICU survivors, i.e. post-ICU acquired weakness. They found it difficult to measure their oxygen saturation using the small button on the side of the ScanWatch. In the ICU Recover Box 2.0, we address these issues by replacing the old devices with a new device: the Corsano CardioWatch 287-2. The CardioWatch seems to be ergonomically better, as there is no need for manual measurements with small buttons. By replacing the Withings devices with the CE-marked CardioWatch, we expect that the requirement of having to be available 24 hours a day will be lifted and that patients with neuromuscular impairment will no longer have problems with measurements.

The CardioWatch can measure pulse rate, HRV, ECG, SpO2, respiratory rate, blood pressure, temperature, activity and sleep. The watch comes with a blood pressure band for monthly blood pressure calibration. An app allows patients to view their own measurements and send push notifications when it is time for blood pressure calibration, an ECG or when the battery is low. The research team uses an online research portal to monitor the measurements.

#### 2.3.1.5 Data collection

The participants will wear the CardioWatch around their wrist. There are different frequency settings for the clinical measurements, ranging from 1/sec to 1/30min, adjustable per parameter. The watch is connected to an app and the app is connected to the online research portal. The raw data will be stored in the Corsano cloud. Questionnaires are conducted over the phone by one of the researchers. We will manually store the answers on Castor DC, a clinical data platform.

There will be a key file linking the patient's identity (LUMC ID) to a study ID. This file will be stored securely on a server accessible only to the research team. In all other platforms used, such as the Corsano app, the research portal or Castor DC, only the study ID will be used.

We will use several Excel files for various purposes, such as tracking study eligibility based on inclusion criteria, managing the informed consent process, maintaining a key file, and scheduling questionnaire phone calls. These files will be stored securely on a server. In addition, any unplanned contact with participants, in case of data transfer problems to the research portal or problems with the CardioWatch, will be documented in a separate Excel file, also stored on a secure server.

### 2.3.2 Preparation for the ICU Recover Box 2.0 study

#### 2.3.2.1 Objective

The study's preparation involved two main aspects. One was to complete the METC application as the study is subject to the WMO and Article 82 of the MDR. This included writing a study protocol and preparing all the necessary documents. The second was to prepare all the logistics of the study. It is crucial to carefully plan and prepare the logistics of a study. Failure of many clinical trials is due to the lack of a structured, practical and businesslike approach to trial management (77). For example, Sadoon et al. conclude that "clinical trials require complex coordination of tasks and involve many resources that need to be planned and managed carefully" (78). During the first pilot study, the lack of a disciplined and structured approach resulted in eligible patients not being enrolled. This can lead to bias and an inaccurate perception of patients' willingness to use and wear smart technology. In order to successfully conduct the ICU Recover Box 2.0 study, it was critical to carefully plan and prepare for the different aspects of the study such as the process of tracking

patient eligibility, managing the enrolment process, organizing patient information and technology installation, monitoring the patients through the research portal and eventually finalizing the study.

#### 2.3.2.2 Research team

The research team consisted of four members throughout the phase, with one member rotating every ten weeks—this was the Technical Medicine intern (TM2). Additionally, the team consisted of a Technical Medicine graduate (TM3) intern and two intensivists. The two intensivists had the main responsibility for the research. The Technical Medicine interns and graduate assisted the intensivists, tried to relieve them of work by performing tasks and were responsible for organizing and thinking out the logistics of the entire study.

#### 2.3.2.3 Deliverables

Deliverables included a study protocol and all necessary forms required for submitting a clinical trial application to the METC. Details of the required documents can be found in Table B.1 in Appendix B.2.

Furthermore, a system of files and portals has been created for researchers to use once the study starts. The location of these resources, the intended actions for each step and the responsibilities assigned to each individual have been carefully thought out and presented in a visual overview, i.e., a metro map. We created two metro maps, one representing the researchers' pathway and one representing the patients' care pathway. Metro mapping is a service design methodology that can be used to design and optimise care pathways (79). The metro map was originally designed to give patients more insight into their care pathway. A metro map has several layers. The first layer consists of metro station icons that indicate where you are in the process. The second layer is the information layer. The third layer shows who is involved at each stop, and the fourth layer provides context. The last layer represents the patient's experience.

A Standard Operating Procedure (SOP) was created to ensure that everything related to the CardioWatch operated correctly. The SOP included installing the app, connecting the wristband to the app and the research portal, and checking that the data is coming through properly and accurately.

## 2.4 Results

As only the preparations for the ICU Recover Box 2.0 study fell within the scope of this thesis, we will only discuss the outputs of this phase.

### 2.4.1 Medical ethical review commission application

Part of this study was to complete the METC application. Under the supervision of and in collaboration with the intensivists from the research team, we have successfully completed the METC application. We have submitted the application and are currently awaiting approval.

### 2.4.2 Study logistics

We have carefully considered the logistics and arranged all necessities to ensure a successful start to the study. We can divide the deliverables into three parts: a visualization of the study logistics,

the standard operating procedure created to facilitate the handling of the CardioWatch specifically and an overview of the take-away points specific to this study, but also generalizable to other clinical studies.

#### 2.4.2.1 Research and patient trajectories

Two TM2 students created a metro map, one representing the researchers' pathway and one representing the patients' care pathway (80, 81). A section of the research metro map and a section of the patient metro map are shown in Figures B.1 and B.2 in Appendix B, respectively. The first metro map created by the TM2 students was a variation on the original concept. Instead of being designed for the patient, it was created for the researchers to provide a better insight into the overall logistics of the study. Figure B.1 shows that the layers have been slightly modified for this purpose. Figure B.2 shows that the focus of the metro map is on the patient experience. The metro maps include the logistics from the time the patient is admitted to the ICU until the end of the study follow-up.

#### 2.4.2.2 Standard operating procedure

The third TM2 student worked on creating an SOP to ensure that everything related to the CardioWatch was well thought out and tested, so there would be no surprises during the study. Throughout this process, many small details emerged that we would not normally have discovered until the study itself had already begun. For example, there are several settings that researchers need to configure in the research portal in advance. The measurement frequency, as well as the parameters and tabs that are visible to the patients in the app, must be set. When the CardioWatch is to be paired with the app on the participants' phone, the wristband must be connected to the charger. This is a small detail, but it is important to ensure that there is always a power outlet available at the installation site. Besides, each patient must set a password. By talking to colleagues who also work with the CardioWatch, we discovered that the app sometimes logs out automatically and that patients who have forgotten their password are unable to log in again. As a result of this experience, we have decided to create and store the passwords for the patients. During testing, by wearing the wristband ourselves, we discovered that many updates are carried out by Corsano and need the patient's approval before they can be installed. If the updates are not carried out, data may not be transmitted correctly. We also found that the phone's battery drains relatively quickly when the Corsano app is constantly fetching data from the CardioWatch. All of these factors must be considered and communicated clearly to the patient. A complete overview of all important details regarding the CardioWatch and how to handle them is outlined in the SOP. As this SOP was created by one of the TM2 students and is not a direct result of this thesis, it has not been included in the Appendix.

#### 2.4.2.3 Take-away points

Several concrete take-away points have emerged during the preparation of the study, which we will briefly discuss.

- Preparing a clinical study is an iterative process. During our preparation, there was a role within the research team that rotated every ten weeks. With each change and a new perspective on the process, new things came to light. This showed that a fresh perspective could contribute positively to the preparation. In addition, it was good to realize that a study can never be prepared to perfection.

- It is important not to lose sight of the objective of the study, which in our case, is to investigate the feasibility of monitoring ICU survivors after discharge. We will primarily determine whether the data are being received correctly and whether the values are realistic. However, we were confronted with an ethical issue because, although it is supposed to be a feasibility study, medical professionals cannot ignore the non-physiological values of the measurements. Therefore, we stated in the protocol that if we notice values that raise concerns about the patient's health, we will inform them about this. However, by adding this action, the goal shifts towards monitoring the patients for clinical purposes, and we are not there yet as the feasibility needs to be established first. The question is, however, whether it is ethical not to communicate with the patient if a worsening trend is visible. The takeaway point is that it is crucial to keep the study's objective in mind while also acting ethically.
- At a time when artificial intelligence is at its peak and we are doing more with data than ever before, it is very important to be prepared to handle data well. This is especially relevant when dealing with personal and health data. During the preparation phase, we were confronted with this in several ways. For example, we had to develop multiple data management plans. This included considering data ownership, data security and data storage. The takeaway is that the data aspect of future studies needs to be well thought out, often in consultation with experts in areas such as security and legal.

## 2.5 Discussion

We aimed to prepare for clinical study the ICU Recover Box 2.0. As the preparation for such a study is mainly operational, it is difficult to present concrete results. Much of the preparation consisted of planning the logistics. We developed a comprehensive step-by-step plan and two TM2 students created a visual representation of it. The planning of the steps and the whole process was within the scope of this thesis, but the visual representation was their contribution. The results provided a brief overview of what such a visual representation looks like, but the complete overview is not a final product of this thesis. The same applies to the standard operating procedure.

Initially, our aim was to obtain METC approval and carry out part of the study as part of this thesis. However, despite our weekly progress meetings, we were unable to achieve this. This showed how time-consuming it is to prepare for a large clinical study. In order to meet all the requirements, different parties were involved. We needed support from people within the LUMC at various stages, as well as from people at Corsano. One of the aspects that delayed the application was the regulations surrounding data and data sharing. The LUMC has its own rules, and since we were collaborating with a third party, an extensive legal agreement had to be created between the LUMC and Corsano. This data processing agreement specifies what exactly happens with the data and who owns it. Although the reliance on third parties slowed down progress, it proved to be a valuable learning experience for everyone involved, as the combination of different backgrounds and specialities ultimately allowed for accurate study execution.

Although it was not possible to start the study during the scope of this thesis, we did prepare for the study as best as we could. We learned a lot about conducting clinical research and identified three main takeaway points. The first emphasized that conducting research and preparing for it is an iterative process, where new insights will continually emerge. Furthermore, we highlighted that it

is important to carefully consider the objective of your study, which could sometimes involve ethical issues that need to be thought through beforehand. Finally, we concluded that we live in a time where data plays an increasingly important role in many fields, including clinical research, which means that handling data, in terms of security, processing and storage, must be well thought out in advance.

## 2.6 Conclusion

The ICU Recover Box 2.0 study is a follow-up to pilot 1.0, in which ICU survivors were monitored using technology after discharge from the ICU. The aim of the follow-up study is to determine the feasibility of home monitoring. Compared to the first pilot study, the most significant change in the second study will be the use of a different device, the Corsano CardioWatch, and the participation of more patients. During the preparations, the entire logistical process was planned out, while simultaneously ensuring everything was in order for the METC application. Once the METC grants approval, the study can start well-prepared.

# References


[1] Marshall JC, Bosco L, Adhikari NK, Connolly B, Diaz JV, Dorman T, et al. What is an intensive care unit? A report of the task force of the World Federation of Societies of Intensive and Critical Care Medicine [Journal Article]. Journal of Critical Care. 2017;37:270-6. Available from: https://www.sciencedirect.com/science/article/pii/S0883944116302404.

[2] Medicine ESoIC. What is Intensive Care? [Web Page];. Available from: https://www.esicm.org/patient-and-family/what-is-intensive-care/.

[3] Harvey MA, Davidson JE. Postintensive Care Syndrome: Right Care, Right Now. . . and Later [Journal Article]. Crit Care Med. 2016;44(2):381-5. 1530-0293 Harvey, Maurene A Davidson, Judy E Journal Article United States 2016/01/16 Crit Care Med. 2016 Feb;44(2):381-5. doi: 10.1097/CCM.0000000000001531.

[4] Rawal G, Yadav S, Kumar R. Post-intensive Care Syndrome: an Overview [Journal Article]. J Transl Int Med. 2017;5(2):90-2. 2224-4018 Rawal, Gautam Yadav, Sankalp Kumar, Raj Journal Article Poland 2017/07/20 J Transl Int Med. 2017 Jun 30;5(2):90-92. doi: 10.1515/jtim-2016-0016. eCollection 2017 Jun.

[5] Vincent JL, Singer M. Critical care: advances and future perspectives [Journal Article]. Lancet. 2010;376(9749):1354-61. 1474-547x Vincent, Jean-Louis Singer, Mervyn Department of Health/United Kingdom Journal Article Research Support, Non-U.S. Gov't England 2010/10/12 Lancet. 2010 Oct 16;376(9749):1354-61. doi: 10.1016/S0140-6736(10)60575-2. Epub 2010 Oct 11.

[6] Gonçalves-Pereira J, Oliveira A, Vieira T, Rodrigues AR, Pinto MJ, Pipa S, et al. Critically ill patient mortality by age: long-term follow-up (CIMbA-LT) [Journal Article]. Ann Intensive Care. 2023;13(1):7. 2110-5820 Gonçalves-Pereira, João Orcid: 0000-0002-7538-3777 Oliveira, André Vieira, Tatiana Rodrigues, Ana Rita Pinto, Maria João Pipa, Sara Martinho, Ana Ribeiro, Sofia Paiva, José-Artur Journal Article Germany 2023/02/11 Ann Intensive Care. 2023 Feb 11;13(1):7. doi: 10.1186/s13613-023-01102-3.

[7] Choi MH, Kim D, Choi EJ, Jung YJ, Choi YJ, Cho JH, et al. Mortality prediction of patients in intensive care units using machine learning algorithms based on electronic health records [Journal Article]. Scientific Reports. 2022;12(1):7180. Available from: https://doi.org/10.1038/s41598-022-11226-4.

[8] Karmali SN, Sciusco A, May SM, Ackland GL. Heart rate variability in critical care medicine: a systematic review [Journal Article]. Intensive Care Med Exp. 2017;5(1):33. 2197-425x Karmali, Shamir N Sciusco, Alberto May, Shaun M Ackland, Gareth L RG/14/4/30736/BHF_/British Heart Foundation/United Kingdom Journal Article Germany 2017/07/14 Intensive Care Med Exp. 2017 Dec;5(1):33. doi: 10.1186/s40635-017-0146-1. Epub 2017 Jul 12.

[9] Shaffer F, Ginsberg JP. An Overview of Heart Rate Variability Metrics and Norms [Journal Article]. Front Public Health. 2017;5:258. 2296-2565 Shaffer, Fred Ginsberg, J P Journal Article Review Switzerland 2017/10/17 Front Public Health. 2017 Sep 28;5:258. doi: 10.3389/fpubh.2017.00258. eCollection 2017.

[10] Mazzeo AT, La Monaca E, Di Leo R, Vita G, Santamaria LB. Heart rate variability: a diagnostic and prognostic tool in anesthesia and intensive care [Journal Article]. Acta Anaesthesiologica Scandinavica. 2011;55(7):797-811. Available from: https://doi.org/10.1111/j.1399-6576.2011.02466.x.

[11] Waxenbaum JA VM Reddy V. Anatomy, Autonomic Nervous System. StatPearls [Internet]: Treasure Island (FL): StatPearls Publishing;; 2023. Available from: https://www.ncbi.nlm.nih.gov/books/NBK539845/.

[12] Johnston BW, Barrett-Jolley R, Krige A, Welters ID. Heart rate variability: Measurement and emerging use in critical care medicine [Journal Article]. J Intensive Care Soc. 2020;21(2):148-57. 2057-360x Johnston, Brian W Orcid: 0000-0003-1634-3297 Barrett-Jolley, Richard Krige, Anton Welters, Ingeborg D Journal Article Review England 2020/06/04 J Intensive Care Soc. 2020 May;21(2):148-157. doi: 10.1177/1751143719853744. Epub 2019 Jun 11.

[13] McCraty R, Shaffer F. Heart Rate Variability: New Perspectives on Physiological Mechanisms, Assessment of Self-regulatory Capacity, and Health risk [Journal Article]. Glob Adv Health Med. 2015;4(1):46-61. 2164-9561 McCraty, Rollin Shaffer, Fred Journal Article Review United States 2015/02/20 Glob Adv Health Med. 2015 Jan;4(1):46-61. doi: 10.7453/gahmj.2014.073.

[14] Electrophysiology TFotESoCtNASoP. Heart Rate Variability [Journal Article]. Circulation. 1996;93(5):1043-65. Doi: 10.1161/01.CIR.93.5.1043. Available from: https://doi.org/10.1161/01.CIR.93.5.1043.

[15] Krumholz HM. Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system [Journal Article]. Health Aff (Millwood). 2014;33(7):1163-70. 2694-233x Krumholz, Harlan M U01 HL105270/HL/NHLBI NIH HHS/United States U01 HL105270-04/HL/NHLBI NIH HHS/United States Journal Article Research Support, N.I.H., Extramural United States 2014/07/10 Health Aff (Millwood). 2014 Jul;33(7):1163-70. doi: 10.1377/hlthaff.2014.0053.

[16] Habehh H, Gohel S. Machine Learning in Healthcare [Journal Article]. Curr Genomics. 2021;22(4):291-300. 1875-5488 Habehh, Hafsa Gohel, Suril Journal Article Review United Arab Emirates 2022/03/12 Curr Genomics. 2021 Dec 16;22(4):291-300. doi: 10.2174/1389202922666210705124359.

[17] Núñez Reiz A, Armengol de la Hoz MA, Sánchez García M. Big Data Analysis and Machine Learning in Intensive Care Units [Journal Article]. Med Intensiva (Engl Ed). 2019;43(7):416-26. 2173-5727 Núñez Reiz, A Armengol de la Hoz, M A Sánchez García, M Journal Article Review Spain 2018/12/29 Med Intensiva (Engl Ed). 2019 Oct;43(7):416-426. doi: 10.1016/j.medin.2018.10.007. Epub 2018 Dec 24.

[18] Sanchez-Pinto LN, Luo Y, Churpek MM. Big Data and Data Science in Critical Care [Journal Article]. Chest. 2018;154(5):1239-48. 1931-3543 Sanchez-Pinto, L Nelson Luo, Yuan Churpek, Matthew M K08 HL121080/HL/NHLBI NIH HHS/United States R01 GM123193/GM/NIGMS NIH HHS/United States R21 LM012618/LM/NLM NIH HHS/United States Journal Article Review United States 2018/05/13 Chest. 2018 Nov;154(5):1239-1248. doi: 10.1016/j.chest.2018.04.037. Epub 2018 May 9.

[19] Makowski D, Pham T, Lau ZJ, Brammer JC, Lespinasse F, Pham H, et al. NeuroKit2: A Python toolbox for neurophysiological signal processing [Journal Article]. Behavior Research Methods. 2021;53(4):1689-96. Available from: https://doi.org/10.3758/s13428-020-01516-y.

[20] Pan J, Tompkins WJ. A Real-Time QRS Detection Algorithm [Journal Article]. IEEE Transactions on Biomedical Engineering. 1985;BME-32(3):230-6.

[21] Bota P, Silva R, Carreiras C, Fred A, da Silva HP. BioSPPy: A Python toolbox for physiological signal processing [Journal Article]. SoftwareX. 2024;26:101712. Available from: https://www.sciencedirect.com/science/article/pii/S2352711024000839.

[22] Wen F, He Ft. An efficient method of addressing ectopic beats: new insight into data preprocessing of heart rate variability analysis [Journal Article]. Journal of Zhejiang University SCIENCE B. 2011;12(12):976-82. Available from: https://doi.org/10.1631/jzus.B1000392.

[23] van Wijk RJ, Quinten VM, van Rossum MC, Bouma HR, Ter Maaten JC. Predicting deterioration of patients with early sepsis at the emergency department using continuous heart rate variability analysis: a model-based approach [Journal Article]. Scand J Trauma Resusc Emerg Med. 2023;31(1):15. 1757-7241 van Wijk, Raymond J Orcid: 0000-0002-2106-7699 Quinten, Vincent M van Rossum, Mathilde C Bouma, Hjalmar R Ter Maaten, Jan C Clinical Trial Journal Article England 2023/04/04 Scand J Trauma Resusc Emerg Med. 2023 Apr 1;31(1):15. doi: 10.1186/s13049-023-01078-w.

[24] Rebergen D, Belur Nagaraj S, Bianchi M, van Putten M, Westover MB, Rosenthal E. ADARRI: a novel method to detect spurious R-peaks in the electrocardiogram for heart rate variability analysis in the intensive care unit [Journal Article]. Journal of Clinical Monitoring and Computing. 2018;32.

[25] Gomes P, Margaritoff P, Plácido da Silva H. pyHRV: Development and Evaluation of an Open-Source Python Toolbox for Heart Rate Variability (HRV); 2019.

[26] Bodenes L, N'Guyen QT, Le Mao R, Ferrière N, Pateau V, Lellouche F, et al. Early heart rate variability evaluation enables to predict ICU patients' outcome [Journal Article]. Sci Rep. 2022;12(1):2498. 2045-2322 Bodenes, Laetitia N'Guyen, Quang-Thang Le Mao, Raphaël Ferrière, Nicolas Pateau, Victoire Lellouche, François L'Her, Erwan Clinical Trial Journal Article Research Support, Non-U.S. Gov't England 2022/02/17 Sci Rep. 2022 Feb 15;12(1):2498. doi: 10.1038/s41598-022-06301-9.

[27] Kakde Y, Bawankule S, Mahajan S, Acharya S, Kumar S, Gaidhane A. Heart rate variability as a prognostic marker in critically ill patients [version 1; peer review: 1 approved with reservations, 1 not approved] [Journal Article]. F1000Research. 2023;12(673). Available from: https://f1000research.com/articles/12-673/v1.

[28] Zhang P, Roberts T, Richards B, Haseler LJ. Utilizing heart rate variability to predict ICU patient outcome in traumatic brain injury [Journal Article]. BMC Bioinformatics. 2020;21(Suppl 17):481. 1471-2105 Zhang, Ping Orcid: 0000-0002-3907-1127 Roberts, Tegan Richards, Brent Haseler, Luke J Journal Article England 2020/12/15 BMC Bioinformatics. 2020 Dec 14;21(Suppl 17):481. doi: 10.1186/s12859-020-03814-w.

[29] de Castilho FM, Ribeiro ALP, da Silva JLP, Nobre V, de Sousa MR. Heart rate variability as predictor of mortality in sepsis: A prospective cohort study [Journal Article]. PLoS One. 2017;12(6):e0180060. 1932-6203 de Castilho, Fábio M Orcid: 0000-0002-0489-3737 Ribeiro, Antonio Luiz P da Silva, José Luiz P Nobre, Vandack de Sousa, Marcos R Journal Article United States 2017/06/28 PLoS One. 2017 Jun 27;12(6):e0180060. doi: 10.1371/journal.pone.0180060. eCollection 2017.

[30] Winchell RJ, Hoyt DB. Spectral analysis of heart rate variability in the ICU: a measure of autonomic function [Journal Article]. J Surg Res. 1996;63(1):11-6. Winchell, R J Hoyt, D B Comparative Study Journal Article United States 1996/06/01 J Surg Res. 1996 Jun;63(1):11-6. doi: 10.1006/jsre.1996.0214.

[31] Adam J, Rupprecht S, Künstler ECS, Hoyer D. Heart rate variability as a marker and predictor of inflammation, nosocomial infection, and sepsis x2013; A systematic review [Journal Article]. Autonomic Neuroscience: Basic and Clinical. 2023;249. Doi: 10.1016/j.autneu.2023.103116. Available from: https://doi.org/10.1016/j.autneu.2023.103116.

[32] Chen WL, Kuo CD. Characteristics of heart rate variability can predict impending septic shock in emergency department patients with sepsis [Journal Article]. Acad Emerg Med. 2007;14(5):392-7. 1553-2712 Chen, Wei-Lung Kuo, Cheng-Deng Journal Article Research Support, Non-U.S. Gov't United States 2007/03/29 Acad Emerg Med. 2007 May;14(5):392-7. doi: 10.1197/j.aem.2006.12.015. Epub 2007 Mar 26.

[33] Chen WL, Chen JH, Huang CC, Kuo CD, Huang CI, Lee LS. Heart rate variability measures as predictors of in-hospital mortality in ED patients with sepsis [Journal Article]. The American Journal of Emergency Medicine. 2008;26(4):395-401. Available from: https://www.sciencedirect.com/science/article/pii/S0735675707004366.

[34] Pontet J, Contreras P, Curbelo A, Medina J, Noveri S, Bentancourt S, et al. Heart rate variability as early marker of multiple organ dysfunction syndrome in septic patients [Journal Article]. Journal of Critical Care. 2003;18(3):156-63. Available from: https://www.sciencedirect.com/science/article/pii/S0883944103000777.

[35] Samsudin MI, Liu N, Prabhakar SM, Chong SL, Kit Lye W, Koh ZX, et al. A novel heart rate variability based risk prediction model for septic patients presenting to the emergency department [Journal Article]. Medicine (Baltimore). 2018;97(23):e10866. 1536-5964 Samsudin, Mas'uud Ibnu Liu, Nan Prabhakar, Sumanth Madhusudan Chong, Shu-Ling Kit Lye, Weng Koh, Zhi Xiong Guo, Dagang Rajesh, R Ho, Andrew Fu Wah Ong, Marcus Eng Hock Journal Article Observational Study United States 2018/06/08 Medicine (Baltimore). 2018 Jun;97(23):e10866. doi: 10.1097/MD.0000000000010866.

[36] Luo X, Gao H, Yu X, Jiang Z, Yang W. Spectral analysis of heart rate variability for trauma outcome prediction: an analysis of 210 ICU multiple trauma patients [Journal Article]. European Journal of Trauma and Emergency Surgery. 2021;47(1):153-60. Available from: https://doi.org/10.1007/s00068-019-01175-5.

[37] Laborde S, Mosley E, Thayer JF. Heart Rate Variability and Cardiac Vagal Tone in Psychophysiological Research - Recommendations for Experiment Planning, Data Analysis, and Data Reporting [Journal Article]. Front Psychol. 2017;8:213. 1664-1078 Laborde, Sylvain Mosley,

Emma Thayer, Julian F Journal Article Review Switzerland 2017/03/08 Front Psychol. 2017 Feb 20;8:213. doi: 10.3389/fpsyg.2017.00213. eCollection 2017.

[38] scikit-learn org. OneHotEncoder [Web Page];. Available from: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html.

[39] Jarapala KN. Categorical Data Encoding Techniques [Blog]; 2023. Available from: https://medium.com/aiskunks/categorical-data-encoding-techniques-d6296697a40f.

[40] Saxena S. What are Categorical Data Encoding Methods | Binary Encoding [Blog]; 2024. Available from: https://www.analyticsvidhya.com/blog/2020/08/types-of-categorical-data-encoding/.

[41] Kunikullaya K, Kunnavil R, Muradi V, Goturu J, Prakash Vs, Murthy S. Normative data and gender differences in heart rate variability in the healthy young individuals aged 18–30 years, a South Indian cross-sectional study [Journal Article]. Indian Pacing and Electrophysiology Journal. 2021;21.

[42] Kleiger RE, Stein PK, Bigger J J T. Heart rate variability: measurement and clinical utility [Journal Article]. Ann Noninvasive Electrocardiol. 2005;10(1):88-101. 1542-474x Kleiger, Robert E Stein, Phyllis K Bigger, J Thomas Jr Journal Article Review United States 2005/01/15 Ann Noninvasive Electrocardiol. 2005 Jan;10(1):88-101. doi: 10.1111/j.1542-474X.2005.10101.x.

[43] Shaffer F, McCraty R, Zerr CL. A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability [Journal Article]. Front Psychol. 2014;5:1040. 1664-1078 Shaffer, Fred McCraty, Rollin Zerr, Christopher L Journal Article Review Switzerland 2014/10/18 Front Psychol. 2014 Sep 30;5:1040. doi: 10.3389/fpsyg.2014.01040. eCollection 2014.

[44] Chan JY, Leow SM, Bea KT, Cheng WK, Phoong SW, Hong ZW, et al.. Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review [Electronic Article]; 2022.

[45] Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in medical research [Journal Article]. Malawi Med J. 2012;24(3):69-71. 1995-7270 Mukaka, M M Journal Article Malawi 2013/05/03 Malawi Med J. 2012 Sep;24(3):69-71.

[46] scikit-learn org. DummyRegressor [Web Page];. Available from: https://scikit-learn.org/stable/modules/generated/sklearn.dummy.DummyRegressor.html.

[47] scikit-learn org. Permutation feature importance [Web Page];. Available from: https://scikit-learn.org/stable/modules/permutation_importance.html.

[48] Çorbacıoğlu Ş K, Aksel G. Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value [Journal Article]. Turk J Emerg Med. 2023;23(4):195-8. 2452-2473 Çorbacıoğlu, Şeref Kerem Orcid: 0000-0001-7802-8087 Aksel, Gökhan Orcid: 0000-0002-5580-3201 Journal Article India 2023/11/29 Turk J Emerg Med. 2023 Oct 3;23(4):195-198. doi: 10.4103/tjem.tjem_182_23. eCollection 2023 Oct-Dec.

[49] Reyes del Paso GA, Langewitz W, Mulder LJM, van Roon A, Duschek S. The utility of low frequency heart rate variability as an index of sympathetic cardiac tone: A review with emphasis

on a reanalysis of previous studies [Journal Article]. Psychophysiology. 2013;50(5):477-87. Available from: https://doi.org/10.1111/psyp.12027.

[50] Miki Y, Suzuki Y, Watanabe E, Hayano J, Yamamoto Y, Nomura T, et al. Long-range correlations in amplitude variability of HF and LF components of heart rate variability. In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC);. p. 6218-21.

[51] Usui H, Nishida Y. The very low-frequency band of heart rate variability represents the slow recovery component after a mental stress task [Journal Article]. PLoS One. 2017;12(8):e0182611. 1932-6203 Usui, Harunobu Orcid: 0000-0002-8178-3355 Nishida, Yusuke Journal Article United States 2017/08/15 PLoS One. 2017 Aug 14;12(8):e0182611. doi: 10.1371/journal.pone.0182611. eCollection 2017.

[52] de Hond AAH, Steyerberg EW, van Calster B. Interpreting area under the receiver operating characteristic curve [Journal Article]. The Lancet Digital Health. 2022;4(12):e853-5. Doi: 10.1016/S2589-7500(22)00188-1. Available from: https://doi.org/10.1016/S2589-7500(22)00188-1.

[53] Kanstrén T. A Look at Precision, Recall, and F1-Score [Web Page]; 2020. Available from: https://towardsdatascience.com/a-look-at-precision-recall-and-f1-score-36b5fd0dd3ec.

[54] Thölke P, Mantilla-Ramos YJ, Abdelhedi H, Maschke C, Dehgan A, Harel Y, et al. Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data [Journal Article]. NeuroImage. 2023;277:120253. Available from: https://www.sciencedirect.com/science/article/pii/S1053811923004044.

[55] Chugh A. MAE, MSE, RMSE, Coefficient of Determination, Adjusted R Squared — Which Metric is Better? [Web Page]; 2020. Available from: https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd032

[56] Jarczok MN, Weimer K, Braun C, Williams DP, Thayer JF, Gündel HO, et al. Heart rate variability in the prediction of mortality: A systematic review and meta-analysis of healthy and patient populations [Journal Article]. Neuroscience Biobehavioral Reviews. 2022;143:104907. Available from: https://www.sciencedirect.com/science/article/pii/S0149763422003967.

[57] Chen IC, Kor CT, Lin CH, Kuo J, Tsai JZ, Ko WJ, et al. High-frequency power of heart rate variability can predict the outcome of thoracic surgical patients with acute respiratory distress syndrome on admission to the intensive care unit: a prospective, single-centric, case-controlled study [Journal Article]. BMC Anesthesiol. 2018;18(1):34. 1471-2253 Chen, I-Chen Kor, Chew-Teng Lin, Ching-Hsiung Kuo, Jane Tsai, Jang-Zern Ko, Wen-Je Kuo, Cheng-Deng Orcid: 0000-0002-3989-1950 Journal Article Research Support, Non-U.S. Gov't England 2018/04/04 BMC Anesthesiol. 2018 Apr 2;18(1):34. doi: 10.1186/s12871-018-0497-5.

[58] Swearingen A, Schafer J, De Lurgio D, Stone J. SDANN as a Predictor of Heart Failure Hospitalizations in Cardiac Resynchronization Therapy Patients [Journal Article]. Journal of Cardiac Failure. 2006;12(6, Supplement):S112-3. Available from: https://www.sciencedirect.com/science/article/pii/S1071916406006695.

[59] Di Franco M, Paradiso M, Riccieri V, Basili S, Mammarella A, Valesini G. Autonomic dysfunction and microvascular damage in systemic sclerosis [Journal Article]. Clinical Rheumatology. 2007;26(8):1278-83. Available from: https://doi.org/10.1007/s10067-006-0492-y.

[60] Hempel L, Sadeghi S, Kirsten T. Prediction of Intensive Care Unit Length of Stay in the MIMIC-IV Dataset [Electronic Article]; 2023.

[61] Baker SG, Schuit E, Steyerberg EW, Pencina MJ, Vickers A, Moons KG, et al. How to interpret a small increase in AUC with an additional risk prediction marker: decision analysis comes through [Journal Article]. Stat Med. 2014;33(22):3946-59. 1097-0258 Baker, Stuart G Schuit, Ewoud Steyerberg, Ewout W Pencina, Michael J Vickers, Andrew Moons, Karel G M Mol, Ben W J Lindeman, Karen S P30 CA008748/CA/NCI NIH HHS/United States Z99 CA999999/Intramural NIH HHS/United States Journal Article England 2014/05/16 Stat Med. 2014 Sep 28;33(22):3946-59. doi: 10.1002/sim.6195. Epub 2014 May 13.

[62] Pencina MJ, D'Agostino RB, Massaro JM. Understanding increments in model performance metrics [Journal Article]. Lifetime Data Anal. 2013;19(2):202-18. 1572-9249 Pencina, Michael J D'Agostino, Ralph B Massaro, Joseph M N01-HC-25195/HC/NHLBI NIH HHS/United States N01 HC025195/HC/NHLBI NIH HHS/United States RC1 HL101056/HL/NHLBI NIH HHS/United States RC1HL101056/HL/NHLBI NIH HHS/United States N01HC25195/HL/NHLBI NIH HHS/United States Journal Article Research Support, N.I.H., Extramural United States 2012/12/18 Lifetime Data Anal. 2013 Apr;19(2):202-18. doi: 10.1007/s10985-012-9238-0. Epub 2012 Dec 16.

[63] Domingos P. A few useful things to know about machine learning [Journal Article]. Commun ACM. 2012;55(10):78–87. Available from: https://doi.org/10.1145/2347736.2347755.

[64] Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part I: basic concepts and first analyses [Journal Article]. Br J Cancer. 2003;89(2):232-8. 1532-1827 Clark, T G Bradburn, M J Love, S B Altman, D G Journal Article Research Support, Non-U.S. Gov't Review England 2003/07/17 Br J Cancer. 2003 Jul 21;89(2):232-8. doi: 10.1038/sj.bjc.6601118.

[65] Needham DM, Davidson J, Cohen H, Hopkins RO, Weinert C, Wunsch H, et al. Improving long-term outcomes after discharge from intensive care unit: report from a stakeholders' conference [Journal Article]. Crit Care Med. 2012;40(2):502-9. 1530-0293 Needham, Dale M Davidson, Judy Cohen, Henry Hopkins, Ramona O Weinert, Craig Wunsch, Hannah Zawistowski, Christine Bemis-Dougherty, Anita Berney, Susan C Bienvenu, O Joseph Brady, Susan L Brodsky, Martin B Denehy, Linda Elliott, Doug Flatley, Carl Harabin, Andrea L Jones, Christina Louis, Deborah Meltzer, Wendy Muldoon, Sean R Palmer, Jeffrey B Perme, Christiane Robinson, Marla Schmidt, David M Scruth, Elizabeth Spill, Gayle R Storey, C Porter Render, Marta Votto, John Harvey, Maurene A Comparative Study Journal Article Review United States 2011/09/29 Crit Care Med. 2012 Feb;40(2):502-9. doi: 10.1097/CCM.0b013e318232da75.

[66] Jensen JF, Thomsen T, Overgaard D, Bestle MH, Christensen D, Egerod I. Impact of follow-up consultations for ICU survivors on post-ICU syndrome: a systematic review and meta-analysis [Journal Article]. Intensive Care Med. 2015;41(5):763-75. 1432-1238 Jensen, J F Thomsen, T Overgaard, D Bestle, M H Christensen, D Egerod, I Journal Article Meta-Analysis Research Support, Non-U.S. Gov't Review Systematic Review United States 2015/03/04 Intensive Care Med. 2015 May;41(5):763-75. doi: 10.1007/s00134-015-3689-1. Epub 2015 Mar 3.

[67] Yanagi N, Kamiya K, Hamazaki N, Matsuzawa R, Nozaki K, Ichikawa T, et al. Post-intensive care syndrome as a predictor of mortality in patients with critical illness: A cohort study [Journal Article]. PLoS One. 2021;16(3):e0244564. 1932-6203 Yanagi, Naoya Kamiya, Kentaro Orcid: 0000-0002-7706-9511 Hamazaki, Nobuaki Matsuzawa, Ryota Nozaki, Kohei Ichikawa, Takafumi Valley, Thomas S Nakamura, Takeshi Yamashita, Masashi Maekawa, Emi Koike, Tomotaka Yamaoka-Tojo, Minako Arai, Masayasu Matsunaga, Atsuhiko Ako, Junya K23 HL140165/HL/NHLBI NIH HHS/United States Journal Article United States 2021/03/11 PLoS One. 2021 Mar 10;16(3):e0244564. doi: 10.1371/journal.pone.0244564. eCollection 2021.

[68] Colbenson GA, Johnson A, Wilson ME. Post-intensive care syndrome: impact, prevention, and management [Journal Article]. Breathe (Sheff). 2019;15(2):98-101. 2073-4735 Colbenson, Gretchen A Johnson, Annie Wilson, Michael E Editorial England 2019/06/14 Breathe (Sheff). 2019 Jun;15(2):98-101. doi: 10.1183/20734735.0013-2019.

[69] McPeake J, Bateson M, Christie F, Robinson C, Cannon P, Mikkelsen M, et al. Hospital readmission after critical care survival: a systematic review and meta-analysis [Journal Article]. Anaesthesia. 2022;77(4):475-85. Available from: https://doi.org/10.1111/anae.15644.

[70] Prescott HC, Langa KM, Iwashyna TJ. Readmission diagnoses after hospitalization for severe sepsis and other acute medical conditions [Journal Article]. Jama. 2015;313(10):1055-7. 1538-3598 Prescott, Hallie C Langa, Kenneth M Iwashyna, Theodore J T32 HL007749/HL/NHLBI NIH HHS/United States U01 AG009740/AG/NIA NIH HHS/United States R01 AG030155/AG/NIA NIH HHS/United States R21 AG044752/AG/NIA NIH HHS/United States P30 AG024824/AG/NIA NIH HHS/United States U01 AG09740/AG/NIA NIH HHS/United States R01 AG0030155/AG/NIA NIH HHS/United States Journal Article Research Support, N.I.H., Extramural Research Support, U.S. Gov't, Non-P.H.S. United States 2015/03/11 JAMA. 2015 Mar 10;313(10):1055-7. doi: 10.1001/jama.2015.1410.

[71] Connect I. IC-nazorg [Web Page];. Available from: https://icconnect.nl/na-de-ic/ic-nazorg/.

[72] Teixeira C, Rosa RG. Post-intensive care outpatient clinic: is it feasible and effective? A literature review [Journal Article]. Rev Bras Ter Intensiva. 2018;30(1):98-111. 1982-4335 Teixeira, Cassiano Rosa, Regis Goulart Journal Article Review Brazil 2018/05/10 Rev Bras Ter Intensiva. 2018 Mar;30(1):98-111. doi: 10.5935/0103-507x.20180016.

[73] Viderman D, Seri E, Aubakirova M, Abdildin Y, Badenes R, Bilotta F. Remote Monitoring of Chronic Critically Ill Patients after Hospital Discharge: A Systematic Review [Journal Article]. J Clin Med. 2022;11(4). 2077-0383 Viderman, Dmitriy Seri, Elena Aubakirova, Mina Abdildin, Yerkin Orcid: 0000-0002-7074-4377 Badenes, Rafael Orcid: 0000-0001-7017-0150 Bilotta, Federico Orcid: 0000-0003-2496-6646 Journal Article Review Switzerland 2022/02/26 J Clin Med. 2022 Feb 15;11(4):1010. doi: 10.3390/jcm11041010.

[74] Treskes RW, van Winden LAM, van Keulen N, van der Velde ET, Beeres S, Atsma DE, et al. Effect of Smartphone-Enabled Health Monitoring Devices vs Regular Follow-up on Blood Pressure Control Among Patients After Myocardial Infarction: A Randomized Clinical Trial [Journal Article]. JAMA Netw Open. 2020;3(4):e202165. 2574-3805 Treskes, Roderick W van Winden, Loes A M van Keulen, Nicole van der Velde, Enno T Beeres, Saskia L M A Atsma, Douwe

E Schalij, Martin Jan Journal Article Randomized Controlled Trial United States 2020/04/17 JAMA Netw Open. 2020 Apr 1;3(4):e202165. doi: 10.1001/jamanetworkopen.2020.2165.

[75] Treskes RW, van den Akker-van Marle ME, van Winden L, van Keulen N, van der Velde ET, Beeres S, et al. The Box-eHealth in the Outpatient Clinic Follow-up of Patients With Acute Myocardial Infarction: Cost-Utility Analysis [Journal Article]. J Med Internet Res. 2022;24(4):e30236. 1438-8871 Treskes, Roderick Willem Orcid: 0000-0003-3210-6005 van den Akker-van Marle, M Elske Orcid: 0000-0002-5269-509x van Winden, Louise Orcid: 0000-0003-2188-0185 van Keulen, Nicole Orcid: 0000-0002-6664-1324 van der Velde, Enno Tjeerd Orcid: 0000-0002-0027-8059 Beeres, Saskia Orcid: 0000-0002-1916-078x Atsma, Douwe Orcid: 0000-0002-9664-2985 Schalij, Martin Jan Orcid: 0000-0003-3767-5231 Journal Article Randomized Controlled Trial Canada 2022/04/26 J Med Internet Res. 2022 Apr 25;24(4):e30236. doi: 10.2196/30236.

[76] Schouten JS. The ICU-Recover Box: Using Smart Technology For Monitoring Health Status After ICU Admission; 2023.

[77] Farrell B, Kenyon S, Shakur H. Managing clinical trials [Journal Article]. Trials. 2010;11(1):78. Available from: https://doi.org/10.1186/1745-6215-11-78.

[78] Sadoon E, Venkatadri U, Ghasemi A. A Conceptual Framework for Logistics Management and Project Planning in the Clinical Trials Industry [Electronic Article]; 2023.

[79] Improving Care with Metro Mapping [Web Page];. Available from: https://metromapping.org/en/#link-service.

[80] Gijben G. ICU-recover Box, Slimme thuismonitoring van post-IC patiënten; 2024.

[81] Wurff Svd. ICU-Recover Box; 2024.

# A

# Appendix

# Appendix A.1

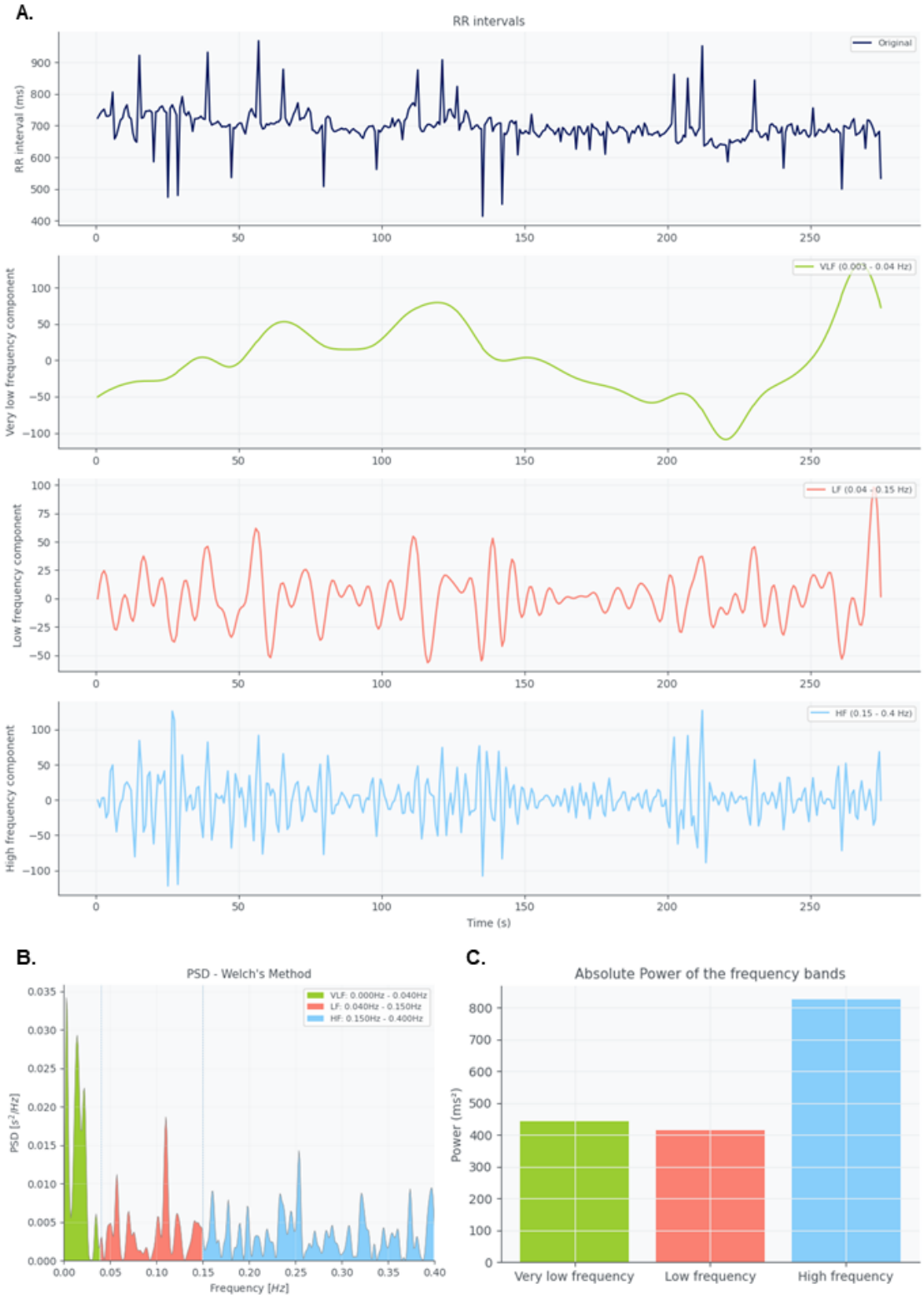



**Figure A.1: Visualisation of the steps from RR intervals to frequency domain parameters for a five minute recording.** A. shows the original RR waveform and the corresponding frequency waveforms for the VLF, LF and HF band, B. shows the power spectra and C. the absolute power in each band.

# Appendix A.2

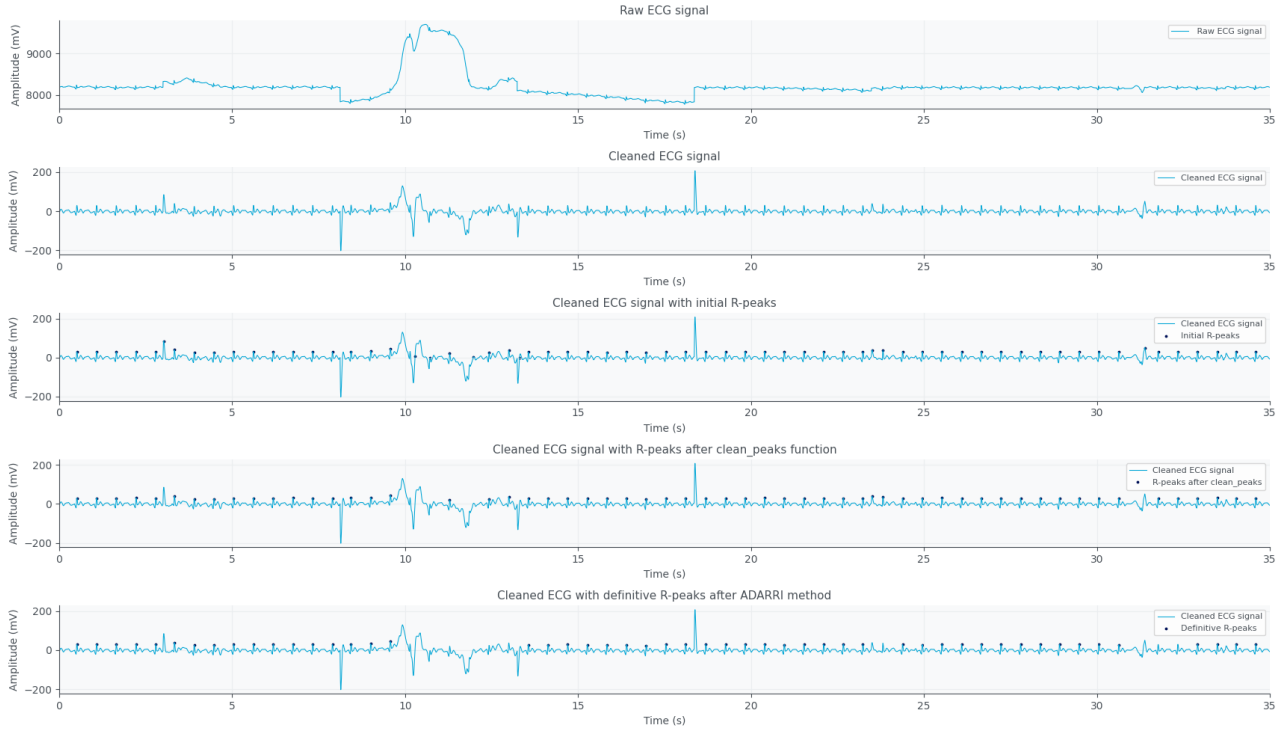



**Figure A.2: A visual step-by-step overview showing how the ECG signal was affected by each processing step.**

# Appendix A.3

**Table A.1: The outcomes of a brief literature review into studies that investigated the predictive value of HRV in a critically ill population (23, 26–36).** APACHE: Acute Physiology and Chronic Health Evaluation, ECG: electrocardiogram, HRV: Heart rate variability, ICU: Intensive care unit, MEWS: Modified Early Warning Score, NEWS: National Early Warning Score, PPG: photoplethysmogram, ROC: receiver operating characteristic, SOFA: Sequential Organ Failure Assessment, TBI: traumatic brain injury.

| Title | Author | Year | N | Objective | Conclusion | Recording setting |
|---|---|---|---|---|---|---|
| Heart rate variability as a marker and predictor of inflammation, nosocomial infection, and sepsis – A systematic review | Adam et al. | 2023 | N/A | To conduct a systematic review of the current literature on the associations and predictive value between HRV and inflammation, and HRV and nosocomial infections/sepsis. | A pro-inflammatory state was associated with reduced total HRV power, affecting both vagal and non-vagal indices. VLF power appears to be most robust in predicting nosocomial infections and sepsis in adults. Promising classical indices include HF (RMSSD), LF, VLF and TP (SDNN). | Not applicable. |
| Early heart rate variability evaluation enables to predict ICU patients' outcome | Bodeness et al. | 2022 | 540 | To identify the most effective indicators of ANS variation for predicting outcomes in ICU patients. | A lower LF/HF ratio, SD2/SD1 ratio, and Shannon entropy values at admission were linked to increased ICU mortality. In multivariate analysis, both LF/HF and Shannon entropy were found to be independently related to mortality. | PPG recordings during two consecutive hours within the first 24 hours of ICU admission. |
| Heart rate variability as predictor of mortality in sepsis: A prospective cohort study | de Castilho et al. | 2017 | 63 | To evaluate HRV as a predictor of 28-day all-cause mortality in septic patients. | In the 20-minute Holter recording, non-survivors had notably lower SDNN, total power, VLF, LF, and LF/HF compared to survivors. An ROC curve for SDNN showed an AUC of 0.772, with SDNN 17 linked to higher mortality risk. The 24-hour Holter HRV parameters did not correlate with 28-day mortality. | 20-minute Holter and a 24-hour Holter on the first day of ICU admission. |

| | | | | | | |
|---|---|---|---|---|---|---|
| Heart rate variability measures as predictors of in-hospital mortality in ED patients with sepsis | Chen et al. | 2008 | 132 | To evaluate how well HRV measurements can predict in-hospital mortality in sepsis patients in the emergency department. | Non-survivors had significantly lower SDNN, TP, VLF power, LF power and LF/HF ratios and higher normalized HF power compared with survivors. Logistic regression identified SDNN and normalized HF power as significant predictors of in-hospital mortality in sepsis patients. | A continuous 10-minute ECG was performed within one hour of arrival at the emergency department. |
| Characteristics of Heart Rate Variability Can Predict Impending Septic Shock in Emergency Department Patients with Sepsis | Chen et al. | 2007 | 81 | To assess whether HRV measurements can predict which septic patients in the emergency department are likely to develop septic shock. | In the septic shock group, LF power, normalized LF power, and the LF/HF ratio were significantly lower compared to the group without septic shock. Multiple logistic regression analysis found RMSSD to be the most reliable predictor of impending septic shock in septic patients in the emergency department. | A continuous 10-minute ECG was performed. |
| Heart rate variability as a prognostic marker in critically ill patients | Kakde et al. | 2023 | 225 | To assess whether HRV can be used as a predictive marker for critically ill patients. | The LF measurement was strongly and independently linked to mortality. A reduction in LF from 24-hour HRV was able to predict mortality with an accuracy of 74%, a specificity of 81.2%, and a sensitivity of 46.7%. | 24-hour holter ECG during ICU admission. |
| Spectral analysis of heart rate variability for trauma outcome prediction: an analysis of 210 ICU multiple trauma patients | Luo et al. | 2019 | 210 | To evaluate and compare short-term spectral HRV indices with widely used trauma scores in predicting multiple trauma outcomes, and to investigate the effectiveness of using them together. | The normalized LF/HF ratio was an independent predictor of 30-day mortality and multiple organ dysfunction syndrome. The combination of normalized LF/HF and conventional trauma scores may improve the accuracy of outcome prediction in multiple trauma. | 5-minute ECG was performed within first 24 hours of ICU admission. |

| | | | | | | |
|---|---|---|---|---|---|---|
| Heart rate variability as early marker of multiple organ dysfunction syndrome in septic patients | Pontet et al. | 2003 | 46 | To assess whether HRV measurements in septic patients without multiple organ dysfunction syndrome can predict which individuals will later develop the syndrome. | SDNN, RMSSD, TINN, LF, and HF were significantly lower in the multiple organ dysfunction syndrome group. Multivariable logistic regression identified LF as the best predictor, with a cut-off point of 18 $ms^2$ established by ROC curves. | A 10-minute ECG was performed during the first 24 hours of admission. |
| A novel heart rate variability based risk prediction model for septic patients presenting to the emergency department | Samsudin et al. | 2018 | 214 | To create a predictive model for evaluating the risk of 30-day in-hospital mortality in septic patients presenting to the emergency department. | The novel risk assessment model, incorporating age, two vital signs, and two HRV parameters (mean NN and DFA 2), performed better than qSOFA, NEWS, and MEWS scores in predicting mortality and adverse events like intubation and ICU admission for septic patients in the emergency department. | 6-minute one-lead ECG. |
| Predicting deterioration of patients with early sepsis at the emergency department using continuous heart rate variability analysis: a model -based approach | van Wijk et al. | 2023 | 168 | To identify key HRV parameters linked to clinical deterioration in early septic patients and to develop a model to compare these parameters with other emergency department scoring systems. | AVNN, ULF, VLF, LF and total power differed between groups (no, stable or progressive organ dysfunction) in the first 12 hours after admission. The predictive accuracy of HRV was similar to other scoring systems, but integration of HRV features into a multivariate model showed potential for predicting progressive organ dysfunction. | ECG waveforms were recorded from arrival at the emergency department up to 48 hours after arrival. HRV parameters were calculated over 5-minute intervals and summarized into 3-hour intervals for analysis. |
| Spectral Analysis of Heart Rate Variability in the ICU: A Measure of Autonomic Function | Winchell et al. | 1996 | 742 | To investigate how changes in HRV affect mortality rates in a surgical ICU. | Low total power HRV and high HF/LF ratio were associated with increased mortality. | HRV measurements were obtained every 6 hours in the ICU. For each HRV measurement, a 5-minute segment of ECG waveform data was acquired. |

| | | | | | | |
|---|---|---|---|---|---|---|
| Utilizing heart rate variability to predict ICU patient outcome in traumatic brain injury | Zhang et al. | 2020 | 26 | To study continuous HRV monitoring during the first 24 hours of ICU stay in severe TBI patients and to develop a predictive outcome system based on HRV data. | HRV-based parameters alone may outperform disease severity scores in predicting outcome in brain injury patients. | ECG signals were collected from bedside monitors throughout the ICU stay, and HRV parameters were calculated over consecutive 30-minute recordings in the time and frequency domains. |

# Appendix A.4

**Table A.2: Included clinical and demographic features.** APACHE: Acute Physiology And Chronic Health Evaluation, BMI: body mass index, bpm: beats per minute (heart rate), breaths per minute (respiratory rate), CPR: cardiopulmonary resuscitation, CRP: C-reactive protein, ICU: Intensive Care Unit, mmHg: millimeter of mercury, SOFA: Sequential Organ Failure Assessment, SpO2: oxygen saturation.

| Feature | Frequency | Data type |
|---|---|---|
| *Bedside monitor* | | |
| Mean heart rate (bpm) | Once for first 24 hours | Float |
| Variance of heart rate | Once for first 24 hours | Float |
| Mean mean arterial blood pressure (mmHg) | Once for first 24 hours | Float |
| Variance of mean arterial blood pressure | Once for first 24 hours | Float |
| Mean systolic arterial blood pressure (mmHg) | Once for first 24 hours | Float |
| Variance of systolic arterial blood pressure | Once for first 24 hours | Float |
| Mean diastolic arterial blood pressure (mmHg) | Once for first 24 hours | Float |
| Variance of diastolic arterial blood pressure | Once for first 24 hours | Float |
| Mean respiratory rate (bpm) | Once for first 24 hours | Float |
| Variance of respiratory rate | Once for first 24 hours | Float |
| Mean SpO2 (%) | Once for first 24 hours | Float |
| Variance of SpO2 | Once for first 24 hours | Float |
| *Electronic health record* | | |
| Age (years) | Once at admission | Integer |
| Gender | Once at admission | Male (M)<br>Female (F) |
| BMI (kg/m²) | Once at admission | Float |
| APACHE IV | Once at admission | Integer |
| APACHE IV diagnosis category | Once at admission | Integer |
| Admission type | Once at admission | Medical admission (1)<br>Emergency surgery (2)<br>Planned surgery (4) |
| Planned admission | Once at admission | Planned (1)<br>Unplanned (0) |
| Highest SOFA score | Once for first 24 hours | Integer |
| Chronic kidney insufficiency<br>or chronic dialysis | Once at admission | Yes (1)<br>No (0) |
| Chronic obstructive pulmonary disease or chronic respiratory insufficiency | Once at admission | Yes (1)<br>No (0) |
| Chronic cardiovascular insufficiency | Once at admission | Yes (1)<br>No (0) |
| Cirrhosis | Once at admission | Yes (1)<br>No (0) |
| Metastasized neoplasm or hematologic malignancy | Once at admission | Yes (1)<br>No (0) |
| Acquired Immunodeficiency Syndrome or immunological insufficiency | Once at admission | Yes (1)<br>No (0) |
| CPR before ICU admission | Once at admission | Yes (1)<br>No (0) |
| Diabetes | Once at admission | Yes (1)<br>No (0) |
| Acute renal failure | Once for first 24 hours | Yes (1)<br>No (0) |
| Mechanical invasive ventilation in the first 24 hours | Once for first 24 hours | Yes (1)<br>No (0) |

| | | |
|---|---|---|
| Highest CRP value (mg/L) | Once for first 24 hours | Float |
| Highest noradrenalin setting (mg) | Once for first 24 hours | Float |
| Dobutamine administration | Once for first 24 hours | Yes (1)<br>No (0) |
| Total dose clonidine and dexmedetomidine (µg) | Once for first 24 hours | Float |
| Total dose propofol (mg) | Once for first 24 hours | Float |
| Total dose midazolam (mg) | Once for first 24 hours | Float |
| Total dose sufentanil (µg) | Once for first 24 hours | Float |
| Total dose labetalol (mg) | Once for first 24 hours | Float |
| Nitroglycerine or nitroprusside administration | Once for first 24 hours | Yes (1)<br>No (0) |
| Total dose furosemide and bumetanide | Once for first 24 hours | Float |
| Highest adrenalin setting (mg) | Once for first 24 hours | Float |
| Oral beta-blockers (metoprolol or sotalol) | Once for first 24 hours | Yes (1)<br>No (0) |
| Total dose amiodarone (mg) | Once for first 24 hours | Float |

## Appendix A.5

During our exploratory data analysis, we manually reviewed the distribution and range of each feature and assessed them to the best of our clinical ability. If there was a question about whether the maximum value was realistic, for example in the case of a medication dose, we checked the medical records using the key file to verify that the value was correct. For the BMI feature, we found a maximum value of 81.3 kg/m². This value was extracted from the NICE database. When we checked the patient's electronic health record, we found a different, more realistic value. This corrected value was used for further analysis. This was the only parameter that was actually incorrect; all other potential outlier values were verified to be correct.

# Appendix A.6

**Table A.3: The possible hyperparameter settings for each model.**

| Model | Possible settings |
|---|---|
| *Classification* | |
| Logistic Regression | Penalty: l1, l2<br>C: 0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000<br>Solver: linlinear, saga |
| Support Vector Machine | Kernel: linear, rbf<br>C: 0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000<br>Gamma (in case of rbf kernel): 0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000 |
| eXtreme Gradient Boosting classifier | Learning rate: 0.01, 0.05, 0.10, 0.20<br>Max depth: 3, 4, 5, 6, 8, 10<br>Min child weight: 1, 3, 5, 7<br>Gamma: 0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000<br>Colsample by tree: 0.3, 0.5, 0.7, 0.9 |
| *Regression* | |
| Lasso Regression | Alpha: 0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000 |
| Support Vector Regression | Kernel: linear, rbf<br>C: 0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000<br>Gamma (in case of rbf kernel): 0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000 |
| eXtreme Gradient Boosting regressor | Learning rate: 0.01, 0.05, 0.10, 0.20<br>Max depth: 3, 4, 5, 6, 8, 10<br>Min child weight: 1, 3, 5, 7<br>Gamma: 0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000<br>Colsample by tree: 0.3, 0.5, 0.7, 0.9 |

# Appendix A.7

**Table A.4: Results from classification models that included the individual HRV measures without clinical features to predict ICU mortality.** AUC: Area under the curve, SVM: Support vector machine, XGBoost: Extreme gradient boosting.

| Feature set | Average AUC over 10 outer folds | Best model | Best hyperparameter settings | Other evaluation metrics |
|---|---|---|---|---|
| *Short-term* | | | | |
| RMSSD<br>Time domain | 0.57 ± 0.11 | SVM | C: 0.0001<br>gamma: 1.0<br>kernel: rbf | Sensitivity: 0.07 ± 0.10<br>Precision: 0.30 ± 0.42<br>F1 score: 0.06 ± 0.10 |
| VLF<br>Frequency domain | 0.53 ± 0.12 | SVM | C: 0.001<br>gamma: 0.001<br>kernel: rbf | Sensitivity: 0.06 ± 0.10<br>Precision: 0.13 ± 0.32<br>F1 score: 0.07 ± 0.12 |
| LF<br>Frequency domain | 0.52 ± 0.12 | SVM | C: 0.1<br>gamma: 0.1<br>kernel: rbf | Sensitivity: 0.01 ± 0.06<br>Precision: 0.00 ± 0.00<br>F1 score: 0.02 ± 0.06 |
| HF<br>Frequency domain | 0.54 ± 0.10 | SVM | C: 0.01<br>gamma: 0.1<br>kernel: rbf | Sensitivity: 0.07 ± 0.10<br>Precision: 0.02 ± 0.05<br>F1 score: 0.09 ± 0.13 |
| LF/HF ratio<br>Frequency domain | 0.46 ± 0.14 | SVM | C: 1.0<br>gamma: 1.0<br>kernel: rbf | Sensitivity: 0.11 ± 0.13<br>Precision: 0.03 ± 0.11<br>F1 score: 0.09 ± 0.10 |
| *Long-term* | | | | |
| SDNN<br>Time domain | 0.55 ± 0.11 | XGBoost | colsample_bytree: 0.3<br>gamma: 0.1<br>learning_rate: 0.01<br>max_depth: 5<br>min_child_weight: 7 | Sensitivity: 0.06 ± 0.14<br>Precision: 0.03 ± 0.08<br>F1 score: 0.07 ± 0.13 |
| SDANN<br>Time domain | 0.52 ± 0.16 | XGBoost | colsample_bytree: 0.3<br>gamma: 0.0001<br>learning_rate: 0.01<br>max_depth: 3<br>min_child_weight: 7 | Sensitivity: 0.15 ± 0.14<br>Precision: 0.16 ± 0.26<br>F1 score: 0.17 ± 0.16 |
| Triangular Index<br>Time domain | 0.54 ± 0.11 | XGBoost | colsample_bytree: 0.3<br>gamma: 0.1<br>learning_rate: 0.01<br>max_depth: 6<br>min_child_weight: 3 | Sensitivity: 0.13 ± 0.13<br>Precision: 0.03 ± 0.11<br>F1 score: 0.14 ± 0.14 |
| ULF<br>Frequency domain | 0.53 ± 0.09 | XGBoost | colsample_bytree: 0.3<br>gamma: 0.0001<br>learning_rate: 0.01<br>max_depth: 5<br>min_child_weight: 5 | Sensitivity: 0.04 ± 0.10<br>Precision: 0.08 ± 0.18<br>F1 score: 0.04 ± 0.12 |

**Table A.5: Results from classification models that included individual HRV measures and combinations of HRV measures in combination with clinical features to predict ICU mortality.** AUC: Area under the curve, HRV: Heart rate variablity, XGBOOST: Extreme gradient boosting.

| Feature set | Average AUC over 10 outer folds | Best model | Best hyperparameter settings | Other evaluation metrics |
|---|---|---|---|---|
| Clinical features only<br>No HRV | 0.72 ± 0.12 | Logistic Regression | C: 0.0001<br>penalty: l2<br>solver: saga | Sensitivity: 0.45 ± 0.13<br>Precision: 0.35 ± 0.47<br>F1 score: 0.39 ± 0.12 |
| All HRV measures<br>VLF, HF, LF/HF ratio, SDANN, Triangular Index, ULF[a] | 0.72 ± 0.12 | XGBoost | colsample_bytree: 0.7<br>gamma: 1.0<br>learning_rate: 0.01<br>max_depth: 8<br>min_child_weight: 1 | Sensitivity: 0.45 ± 0.13<br>Precision: 0.25 ± 0.36<br>F1 score: 0.35 ± 0.14 |
| *Short-term* | | | | |
| RMSSD<br>Time domain | 0.71 ± 0.10 | Logistic Regression | C: 0.0001<br>penalty: l2<br>solver: saga | Sensitivity: 0.44 ± 0.14<br>Precision: 0.33 ± 0.47<br>F1 score: 0.33 ± 0.16 |
| VLF<br>Frequency domain | 0.71 ± 0.12 | XGBoost | colsample_bytree: 0.5<br>gamma: 0.1<br>learning_rate: 0.01<br>max_depth: 10<br>min_child_weight: 1 | Sensitivity: 0.45 ± 0.13<br>Precision: 0.18 ± 0.34<br>F1 score: 0.38 ± 0.11 |
| LF<br>Frequency domain | 0.70 ± 0.09 | Logistic Regression | C: 0.0001<br>penalty: l2<br>solver: saga | Sensitivity: 0.45 ± 0.13<br>Precision: 0.28 ± 0.42<br>F1 score: 0.39 ± 0.12 |
| HF<br>Frequency domain | 0.73 ± 0.09 | XGBoost | colsample_bytree: 0.5<br>gamma: 0.1<br>learning_rate: 0.01<br>max_depth: 8<br>min_child_weight: 1 | Sensitivity: 0.44 ± 0.14<br>Precision: 0.45 ± 0.50<br>F1 score: 0.39 ± 0.12 |
| LF/HF ratio<br>Frequency domain | 0.71 ± 0.08 | Logistic Regression | C: 0.0001<br>penalty: l2<br>solver: saga | Sensitivity: 0.45 ± 0.13<br>Precision: 0.28 ± 0.42<br>F1 score: 0.40 ± 0.12 |
| Short-term measures combined<br>VLF, HF, LF/HF ratio[b] | 0.76 ± 0.10 | XGBoost | colsample_bytree: 0.7<br>gamma: 0.0001<br>learning_rate: 0.01<br>max_depth: 3<br>min_child_weight: 1 | Sensitivity: 0.44 ± 0.14<br>Precision: 0.33 ± 0.34<br>F1 score: 0.33 ± 0.16 |
| *Long-term* | | | | |
| SDNN<br>Time domain | 0.72 ± 0.11 | Logistic Regression | C: 0.0001<br>penalty: l2<br>solver: saga | Sensitivity: 0.44 ± 0.11<br>Precision: 0.23 ± 0.42<br>F1 score: 0.38 ± 0.11 |
| SDANN<br>Time domain | 0.72 ± 0.12 | XGBoost | colsample_bytree: 0.5<br>gamma: 0.0001<br>learning_rate: 0.01<br>max_depth: 10<br>min_child_weight: 1 | Sensitivity: 0.43 ± 0.12<br>Precision: 0.13 ± 0.22<br>F1 score: 0.36 ± 0.13 |
| Triangular Index<br>Time domain | 0.71 ± 0.11 | Logistic Regression | C: 0.0001<br>penalty: l2<br>solver: liblinear | Sensitivity: 0.45 ± 0.13<br>Precision: 0.35 ± 0.46<br>F1 score: 0.39 ± 0.12 |
| ULF<br>Frequency domain | 0.71 ± 0.10 | XGBoost | colsample_bytree: 0.5<br>gamma: 0.01<br>learning_rate: 0.01<br>max_depth: 10<br>min_child_weight: 1 | Sensitivity: 0.45 ± 0.13<br>Precision: 0.27 ± 0.42<br>F1 score: 0.39 ± 0.12 |
| Long-term measures combined<br>SDANN, Triangular Index, ULF[c] | 0.72 ± 0.12 | XGBoost | colsample_bytree: 0.9<br>gamma: 0.01<br>learning_rate: 0.01<br>max_depth: 3<br>min_child_weight: 1 | Sensitivity: 0.45 ± 0.12<br>Precision: 0.19 ± 0.36<br>F1 score: 0.39 ± 0.11 |

[a] SDNN, RMSSD and LF were removed based on their correlation with the other HRV measures
[b] RMSSD and LF were removed based on their correlation with the power in the HF band
[c] SDNN was removed based on its correlation with the SDANN

# Appendix A.8

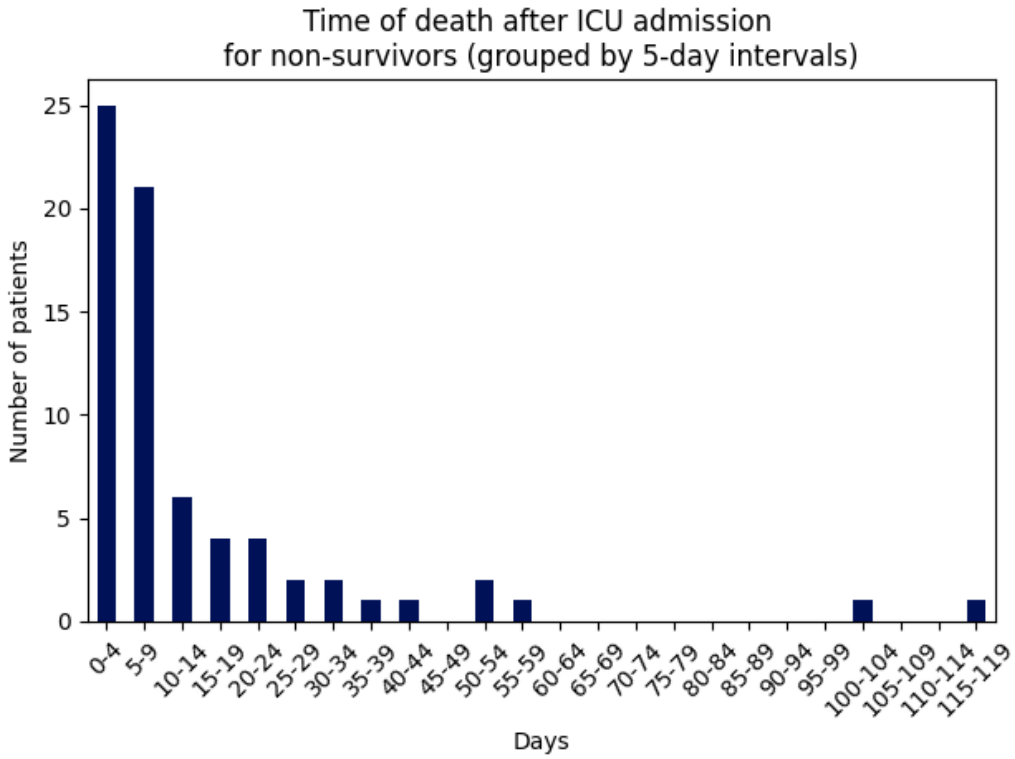



**Figure A.3:** This indicates how long they were admitted before passing away at the ICU.

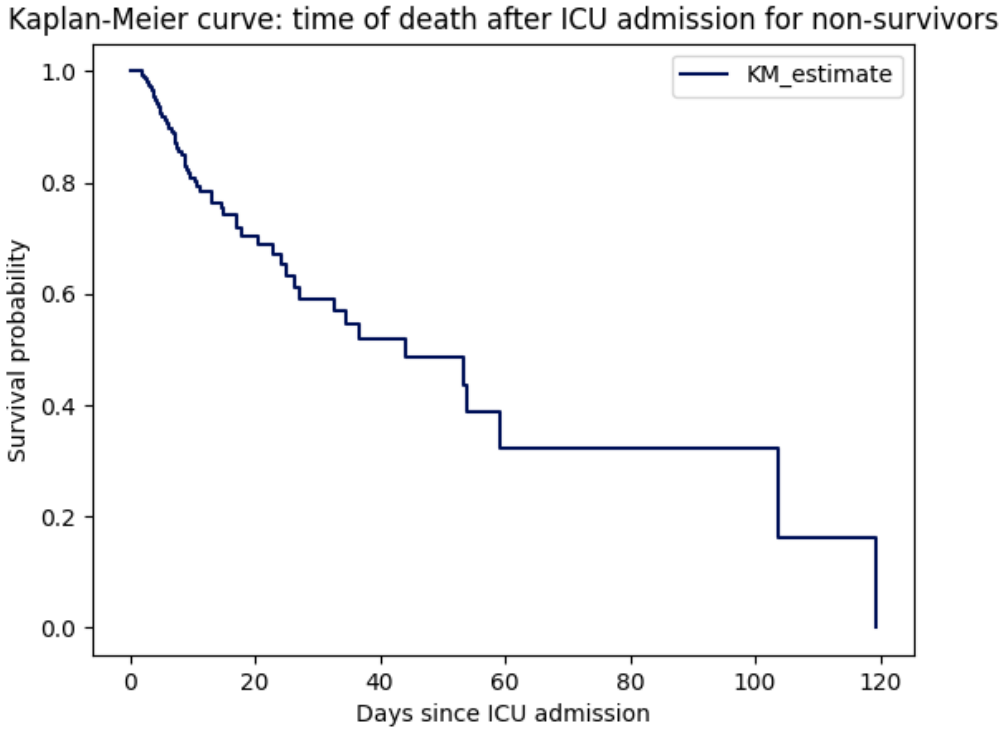



**Figure A.4: Kaplan-Meier curve showing the survival duration of non-ICU survivors.**

# Appendix A.9

**Table A.6: Results from regression models that included the individual HRV measures without clinical features to predict ICU LOS.** MAE: Mean absolute error, RMSE: Root mean squared error, SVR: Support vector regressor, XGBoost: Extreme gradient boosting.

| Feature set | Average MAE over 10 outer folds | Best model | Best hyperparameter settings | Other evaluation metrics |
|---|---|---|---|---|
| *Short-term* | | | | |
| RMSSD<br>Time domain | 5.19 ± 1.22 | SVR | C: 0.01<br>gamma: 100.0<br>kernel: rbf | RMSE: 15.32 ± 16.21<br>$R^2$: -0.04 ± 0.09 |
| VLF<br>Frequency domain | 5.18 ± 1.20 | SVR | C: 1<br>gamma: 100.0<br>kernel: rbf | RMSE: 9.86 ± 3.60<br>$R^2$: -0.04 ± 0.06 |
| LF<br>Frequency domain | 5.17 ± 1.23 | SVR | C: 0.01<br>kernel: linear | RMSE: 9.87 ± 3.61<br>$R^2$: -0.04 ± 0.07 |
| HF<br>Frequency domain | 5.18 ± 1.22 | SVR | C: 0.1<br>gamma: 0.1<br>kernel: rbf | RMSE: 9.91 ± 3.64<br>$R^2$: -0.05 ± 0.09 |
| LF/HF ratio<br>Frequency domain | 5.17 ± 1.23 | SVR | C: 0.01<br>gamma: 100<br>kernel: rbf | RMSE: 9.86 ± 3.60<br>$R^2$: -0.04 ± 0.06 |
| *Long-term* | | | | |
| SDNN<br>Time domain | 5.18 ± 1.23 | SVR | C: 0.1<br>gamma: 100<br>kernel: rbf | RMSE: 10.10 ± 3.26<br>$R^2$: -0.16 ± 0.39 |
| SDANN<br>Time domain | 5.18 ± 1.24 | SVR | C: 0.01<br>gamma: 1000<br>kernel: rbf | RMSE: 9.84 ± 3.60<br>$R^2$: -0.04 ± 0.07 |
| Triangular Index<br>Time domain | 5.20 ± 1.28 | SVR | C: 1000<br>gamma: 1<br>kernel: rbf | RMSE: 9.86 ± 3.50<br>$R^2$: -0.04 ± 0.08 |
| ULF<br>Frequency domain | 5.15 ± 1.21 | SVR | C: 1000<br>gamma: 1<br>kernel: rbf | RMSE: 10.16 ± 3.76<br>$R^2$: -0.10 ± 0.16 |

**Table A.7: Results from regression models that included individual HRV measures and combinations of HRV measures in combination with clinical features to predict ICU LOS.** HRV: Heart rate variablity, MAE: Mean absolute error, RMSE: Root mean squared error, XGBoost: Extreme gradient boosting.

| Feature set | Average MAE over 10 outer folds | Best model | Best hyperparameter settings | Other evaluation metrics |
|---|---|---|---|---|
| Clinical features only<br>No HRV | 5.09 ± 1.17 | SVR | C: 0.01<br>kernel: linear | RMSE: 9.70 ± 3.34<br>$R^2$: -0.04 ± 0.26 |
| All HRV measures<br>VLF, HF, LF/HF ratio, SDANN, Triangular Index, ULF[a] | 5.08 ± 1.17 | SVR | C: 1.0<br>gamma: 0.01<br>kernel: rbf | RMSE: 9.73 ± 3.43<br>$R^2$: -0.03 ± 0.17 |
| *Short-term* | | | | |
| RMSSD<br>Time domain | 5.08 ± 1.17 | SVR | C: 1.0<br>gamma: 0.01<br>kernel: rbf | RMSE: 9.73 ± 3.43<br>$R^2$: -0.03 ± 0.17 |
| VLF<br>Frequency domain | 5.09 ± 1.16 | SVR | C: 1.0<br>gamma: 0.01<br>kernel: rbf | RMSE: 9.58 ± 3.55<br>$R^2$: 0.02 ± 0.13 |
| LF<br>Frequency domain | 5.08 ± 1.18 | SVR | C: 0.01<br>kernel: linear | RMSE: 9.65 ± 3.69<br>$R^2$: 0.03 ± 0.13 |
| HF<br>Frequency domain | 5.10 ± 1.17 | SVR | C: 0.01<br>kernel: linear | RMSE: 9.69 ± 3.56<br>$R^2$: 0.01 ± 0.19 |
| LF/HF ratio<br>Frequency domain | 5.08 ± 1.19 | SVR | C: 0.01<br>kernel: linear | RMSE: 9.53 ± 3.46<br>$R^2$: 0.03 ± 0.13 |
| Short-term measures combined<br>VLF, HF, LF/HF ratio[b] | 5.07 ± 1.17 | SVR | C: 0.01<br>kernel: linear | RMSE: 9.70 ± 3.49<br>$R^2$: -0.01 ± 0.16 |
| *Long-term* | | | | |
| SDNN<br>Time domain | 5.10 ± 1.17 | SVR | C: 1.0<br>gamma: 0.01<br>kernel: rbf | RMSE: 9.55 ± 3.45<br>$R^2$: 0.02 ± 0.13 |
| SDANN<br>Time domain | 5.08 ± 1.19 | SVR | C: 1.0<br>gamma: 0.01<br>kernel: rbf | RMSE: 9.69 ± 3.73<br>$R^2$: 0.01 ± 0.14 |
| Triangular Index<br>Time domain | 5.09 ± 1.18 | SVR | C: 1.0<br>gamma: 0.01<br>kernel: rbf | RMSE: 9.61 ± 3.65<br>$R^2$: 0.01 ± 0.12 |
| ULF<br>Frequency domain | 5.08 ± 1.18 | SVR | C: 0.01<br>kernel: linear | RMSE: 9.95 ± 3.57<br>$R^2$: 0.03 ± 0.14 |
| Long-term measures combined<br>SDANN, Triangular Index, ULF[c] | 5.10 ± 1.19 | SVR | C: 1.0<br>gamma: 0.01<br>kernel: rbf | RMSE: 9.66 ± 3.56<br>$R^2$: 0.01 ± 0.12 |

[a] SDNN, RMSSD and LF were removed based on their correlation with the other HRV measures
[b] RMSSD and LF were removed based on their correlation with the power in the HF band
[c] SDNN was removed based on its correlation with the SDANN

# Appendix A.10

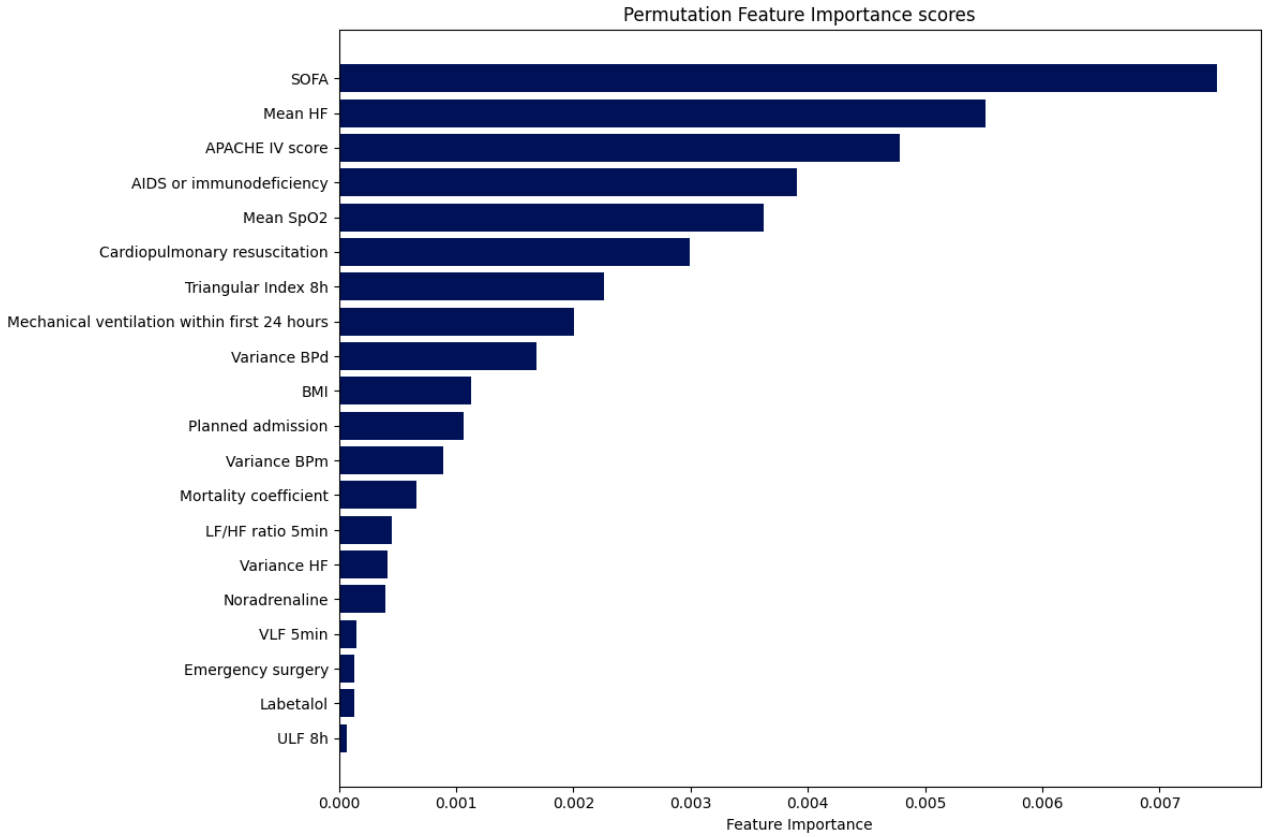



**Figure A.5: Feature importance plot of a support vector regressor combining all clinical features with all HRV measures.**

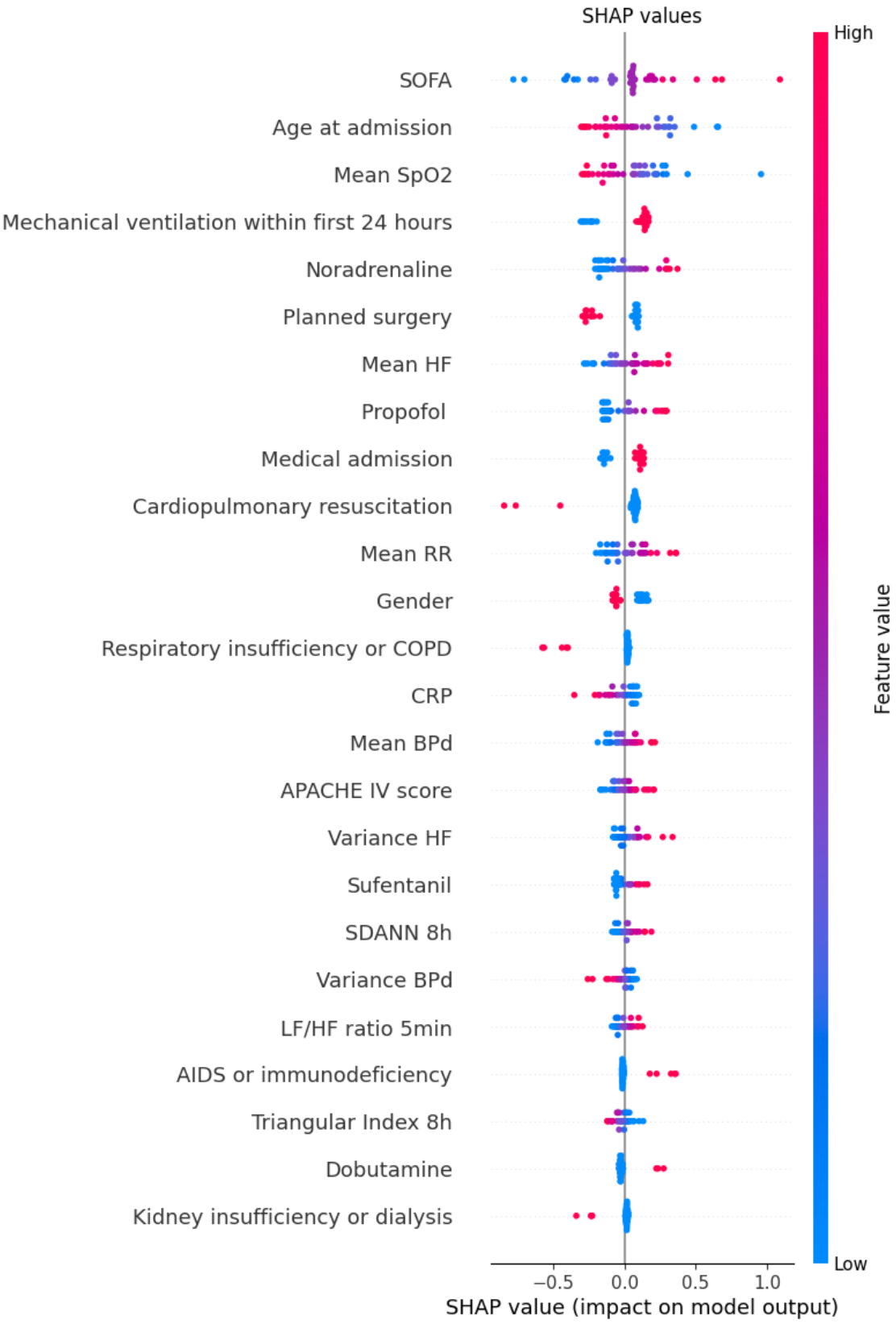



**Figure A.6: SHAP values a support vector regressor combining all clinical features with all HRV measures.**

# B

# Appendix

## Appendix B.1

This page was intentionally left blank.

## Appendix B.2

The following documents are required when submitting a METC application for clinical research on medical devices under the scope of Article 82. The names of these documents will be listed in Dutch.

**A1** Aanbiedingsbrief
**A2** Machtiging van de verrichter[a]
**A2** Machtiging person EU als sponsor niet in EU gevestigd is[a]
**B1** ABR formulier
**B6** CCMO-formulier beëindiging studie[a]
**C1** Onderzoeksprotocol
**C2** Protocolamendementen[a]
**D2** Investigational Medical Device Dossier (IMDD)[b]
**D2** Instruction for use
**D4** Verklaring fabrikant over veiligheid en prestaties medisch hulpmiddel[b]
**E1/E2** Informatiebrief en toestemmingsformulier proefpersonen
**E3** Wervingsmateriaal proefpersonen[a]
**E4** Overig voorlichtingsmateriaal[a]
**E5** Nieuwsbrieven/brieven resultaten[a]
**F1** Vragenlijsten[a]
**F2** Patiëntendagboeken[a]
**F3** Patiëntenkaarten[a]
**F4** Overig[a]
**G1** WMO-proefpersonenverzekering
**G2** Aansprakelijkheidsverzekeringen
**H2** CV coördinerend onderzoeker
**I1** Lijst van deelnemende centra
**I2** Onderzoeksverklaringen of Verklaring Geschiktheid Onderzoeksinstelling (VGO)
**I3** CV's hoofdonderzoekers
**I4** Overige centruminformatie[a]
**J1** Vergoedingen proefpersonen
**J2** Vergoedingen onderzoekers en centra
**K1** Adviezen andere instanties
**K2** Beoordeling andere EU-lidstaten[a]
**K3** Onderzoekscontracten
**K4** Relevante publicaties[a]
**K5** Charter DSMB[a]
**K6** Overige informatie[a]

[a] If applicable.
[b] Only required for non-CE marked medical devices or CE marked medical devices used outside of their intended use.

# Appendix B.3

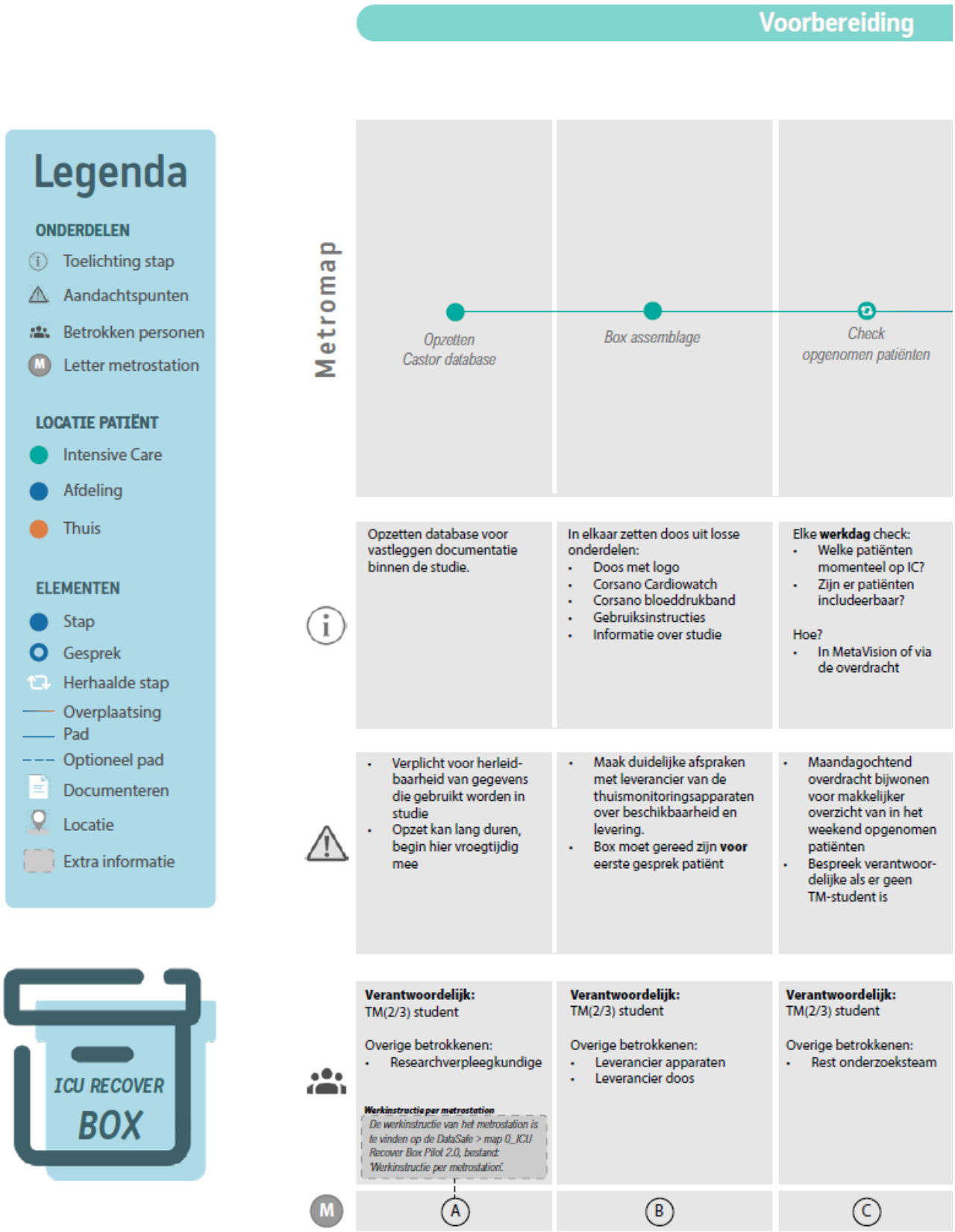



**Figure B.1: A section of the research metro map (80).**

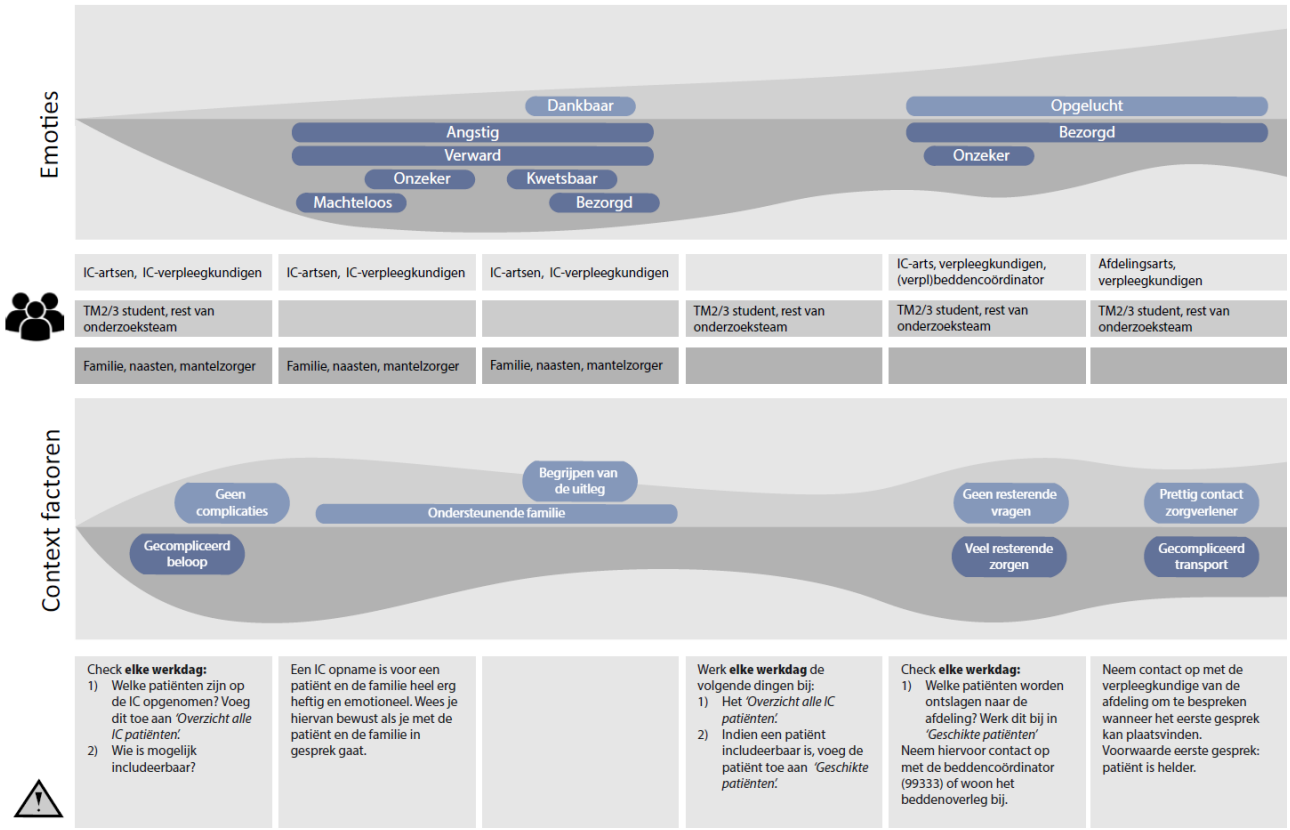



**Figure B.2: A section of the patient metro map (81).**