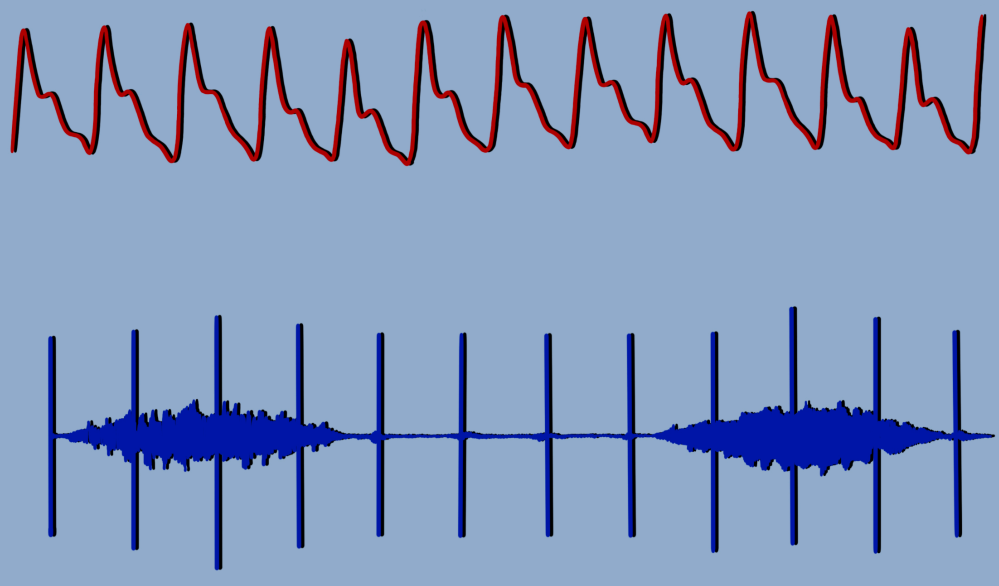
UNIVERSITY OF TWENTE.

PREDICTING DETERIORATION OF EARLY SEPSIS PATIENTS USING WEARABLES

A.D. Schoonhoven



Master Thesis Technical Medicine

Personal information:
A.D. Schoonhoven, BSc
studentnumber: s1663968
a.d.schoonhoven@student.utwente.nl/ annaschoonhoven@hotmail.com

Internship:
University Medical Center Groningen
Hanzeplein 1, 9713 GZ Groningen
Emergency Department
January 2022 - February 2023

Educational institution:
University of Twente
Drienerlolaan 5, 7522 NB Enschede
Faculty Science and Technology
Master Technical Medicine
Track: Medical Sensing and Stimulation

Graduation committee:
Chair: Prof. dr. D. W. Donker
Clinical supervisor: Prof. dr. J.C. ter Maaten
Technical supervisor: M.P. Mulder, MSc
Process supervisor: dr. M. Groenier
Extra member: R.J. van Wijk, MSc
External member: M.M.L.H Verhulst, MSc

Colloquium date:
22 February 2023, 15.00 University of Twente

ACKNOWLEDGEMENTS

In september 2015, nu 7.5 jaar geleden, begon ik aan mijn studie Technische Geneeskunde in Enschede. Een moeilijke keuze was dit niet, aangezien ik op zoek was naar zowel medische als technische uitdaging. Ik heb een fijne tijd gehad in Enschede en daarna voor de M2 stages verhuisd door Nederland, waar ik stage heb gelopen in Utrecht, Nijmegen en Arnhem. Tijdens deze stages realiseerde ik me hoe leuk ik de signaalanalyse vond en dat het werken met Matlab me steeds beter af ging. In de kliniek merkte ik dat ik vooral de acute en levensbedreigde patiënten erg interessant vond. Mijn M3 stage ben ik daarom in januari 2022 begonnen op de Spoedeisende Hulp in het UMCG in Groningen. Hier heb ik een jaar gewerkt aan onderzoek naar het voorspellen van verslechtering van sepsis patiënten op basis van cardiovasculaire en respiratoire signalen. Het schrijven van deze scriptie heb ik echter niet alleen gedaan en daarom wil ik graag een aantal mensen bedanken.

Ten eerste wil ik Jan ter Maaten bedanken voor alle medische input. In jouw drukke agenda was altijd een gaatje te vinden om samen te overleggen. Daarnaast was het fijn om te merken hoe je naast het onderzoek ook altijd geinteresseerd was in hoe het ging met mij als persoon. Raymond van Wijk wil ik bedanken voor de dagelijkse begeleiding en je onuitputbare input. Ondanks je eigen drukke planning en veel studenten die tijd van je vroegen, maakte je duidelijk aan mij dat er altijd tijd was om te brainstormen. Het was fijn hoe jij altijd meteen snapte waar ik het over had en me weer aan het denken kon zetten. Dan wil ik Hjalmar Bouma bedanken voor alle input tijdens research meetings, je positiviteit en het mogelijk maken van dit onderzoek. Paulien Engelbert kan niet missen in dit rijtje, dankjewel voor het opzetten van dit onderzoek en het implementeren van de wearables op de spoedeisende hulp.

Vanuit de Universiteit Twente wil ik Hermie Hermens in de eerste maanden bedanken voor zijn technische bijdragen op het gebied van EMG metingen. Voor de technische begeleiding in de andere maanden wil ik Dirk Donker bedanken voor het kritisch meedenken vanuit een ander perspectief. Marijn Mulder, bedankt voor je input en meedenken, het keer op keer doorlezen van mijn werk en voor het uitdagen om mijn schrijven te verbeteren. Tot slot wil ik Marleen Groenier bedanken voor het begeleiden van mijn professionele ontwikkeling gedurende de afgelopen twee jaar.

Bovendien wil ik graag de hele onderzoeksgroep bedanken voor de gezellige momenten dit jaar tijdens de pauzes, de lange uren op kantoor en de borrels in het weekend. Door jullie enthousiasme ging ik elke dag met veel plezier aan het werk. Ik wil ook mijn huisgenoten bedanken die er het hele jaar door voor mij waren en me altijd wisten op te vrolijken als ik er doorheen zat. Een speciaal dankjewel aan Jaika voor het ontwerp van mijn voorblad.

Als laatst wil ik mijn familie, vrienden en studiegenoten bedanken voor alle steun deze periode. Ik wil specifiek Susanne, Maaike, Cindel, Eva, Manon, Chiara, Stefan, Leen, Lisa, Jorrit en Nicole bedanken voor de samenwerking en de onvergetelijke momenten samen in Enschede. Jullie hebben mijn studietijd tot een fantastische periode gemaakt.

ABSTRACT

Introduction
Sepsis is a severe condition that can lead to septic shock and multiple organ failure if not promptly recognized and treated. Early detection of deterioration is challenging, as continuous monitoring is uncommon in nursing wards. However, the cardiovascular and respiratory system can be monitored through wearables able to capture photoplethysmography (PPG) and diaphragm electromyography (EMG). We hypothesized that an artificial intelligence-based analysis of trends in PPG signals allows accurate prediction of patient deterioration.

Methods
Continuously measured PPG and EMG waveforms were obtained from the Acutelines data-biobank at the University Medical Centre. A Matlab algorithm was used to filter and analyze 12-hour measurements, extract features, and determine trends such as the number of slope changes and overall slope. A Mann-Whitney U test was used to calculate significant differences in trend features between both groups. Five supervised Machine Learning models were developed with five-fold cross-validation.

Results
We included 189 patients with early sepsis, of which 25 deteriorated, defined as death, Intensive Care Unit admission, or a rise in SOFA score by at least 2 points within 48 hours after acute admission. Using PPG trend features selected through statistical evaluation, we demonstrated effective predictive modeling. The K-nearest neighbor (KNN) model displayed good performance: the model's accuracy was 75%, sensitivity 73%, specificity 78%, and an AUROC 0.84. Unfortunately, there was insufficient EMG data from deteriorating patients for inclusion in the model.

Conclusion
This study demonstrates the potential of using PPG trend features for early prediction of patient deterioration. Further research is needed to evaluate the generalization of the models with a separate test set. Yet, the results seem promising to improve patient safety via telemedicine in hospitals and at home.

CONTENTS

LIST OF ABBREVIATIONS

ABP	Arterial Blood Pressure
AC	Alternating Current
AI	Augmentation Index
APG	Acceleration Plethysmography
ARDS	Acute Respiratory Distress Syndrome
AUROC	Area Under the Receiver Operating Characteristics
CNS	Central nervous system
CO	Cardiac Output
CT	Crest Time
DAMP	Damage-associated Molecular Patterns
DC	Direct Current
DPA	Diastolic Peak Amplitude
DT	Delta Time
ECG	Electrocardiography
ED	Emergency Department
EMG	Electromyography
ESI	Emergency Severity Index
FDR	False Discovery Rate
GCS	Glasgow Coma Scale
HR	Heart Rate
HRV	Heart Rate Variability
ICU	Intensive Care Unit
IPA	Infliction Point Area
IQR	Interquartile Ranges
KNN	K-Nearest Neighbors
LR	Logistic Regression
MAP	Mean Arterial Pressure
PI	Pulse Interval
PPG	Photoplethysmography
PtP	Peak-to-Peak
PW	Pulse Width
RA	Relative Amplitude
RI	Reflection Index
ROS	Reactive Oxygen Species
SCCM	Society of Critical Care Medicine
SIRS	Systemic Inflammatory Response Syndrome
SMOTE	Synthetic Minority Oversampling Technique
SNR	Signal-to-Noise Ratio
SPA	Systolic Peak Amplitude
SpO_2	Oxygen Saturation
(q)SOFA	(Quick) Sequential Organ Failure Assessment
SV	Stroke Volume
SVM	Support Vector Machine
TS	Template Subtraction
UMCG	University Medical Centre Groningen
SQI	Signal Quality Index
RF	Random Forest
RMS	Root Mean Square

1 INTRODUCTION

Sepsis is a global healthcare issue and continues to be the leading cause of death from infection [1]. Approximately 49 million people are affected by sepsis every year and it is estimated that 11 million deaths are caused by the syndrome, accounting for up to 19.7% of all deaths worldwide [2]. Sepsis, a potentially life-threatening condition characterized by organ dysfunction resulting from a dysregulated host response to infection, can progress to septic shock and multiple organ failure if not promptly recognized and treated. The mortality rate for septic shock and multiple organ failure is notably high, with a rate of 40% or more reported in Napolitano et al [1]. Early identification and prediction of patients at risk of sepsis is therefore crucial for improving treatment outcomes. [3].

Identifying patients at risk can be difficult, especially in patients who have been discharged or transferred to a standard nursing ward without continuous monitoring. Although these patients may not initially seem to require constant monitoring, a significant portion will deteriorate within 72 hours [4]. While vital signs are checked on nursing wards during regular check-ups by nurses, a change in these signs may go unnoticed for several hours, potentially leading to delays in detecting the deterioration of a patient [5]. Even when a decline in vital signs is detected, it may not be the earliest indication of a patient's worsening condition. For example, decreased oxygen saturation is commonly used to identify respiratory failure, but there may be a delay between the worsening of pneumonia and a decrease in oxygen saturation. In such cases, continuously monitoring the patient's respiratory rate may provide earlier indications of respiratory deterioration. Therefore, it might be beneficial for the patient to be continuously monitored in order to gain insight into the patient's condition and facilitate early detection of deterioration.

Continuous monitoring of patients in nursing wards as is done in the Intensive Care Unit (ICU) is technically not possible. However, cheap and non invasive wearables can be used at general wards for continuous monitoring, as wireless wearable devices have recently become more widely available [6]. To effectively detect deterioration, it is important to monitor the organs that can be continuously measured. During deterioration, changes can occur in various organ systems, such as the cardiovascular, respiratory, hepatic, renal and central nervous system (CNS). It can be difficult to continuously monitor deterioration in the liver and kidneys, as they are often evaluated through blood analyses. However, the cardiovascular and respiratory systems can be continuously monitored through the use of photoplethysmography (PPG) and diaphragm electromyography (EMG) wearables. PPG sensors can measure changes in the cardiovascular system, such as decreased systemic vascular resistance and changes in cardiac output, that may occur during the early stages of sepsis [7]. EMG activity of the diaphragm, the main muscle involved in respiration, can be used to detect changes in the respiratory system, such as an increase in the respiratory rate and a change in the work of breathing.

Continuously measured signals provide a wealth of information. The fact that these signals are measured over an extended period of time allows for the analysis of trends over time. Trend analysis in continuously measured signals is useful, as it enables the identification of subtle changes over time that may not be immediately apparent when considering only the absolute values of measurements. These changes can be indicative of underlying physiological processes and can provide important insights into a patient's condition. In previous research, Churpek et al. showed that adding trends of vital signs significantly increased the accuracy of models designed to detect critical illness on the wards. Vital signs used here were the temperature, heart rate, respiratory rate, oxygen saturation, diastolic blood pressure and systolic blood pressure [8, 9].

Before wearable devices can be utilized to predict deterioration at an early stage, it is necessary to determine the extent to which the trends observed in these devices are associated with deterioration and which combination of trends can be most effectively used in predictive models. Machine learning models are well-suited for this purpose, because they can handle large and complex data sets and can automatically identify patterns and relationships within the data. Therefore, the following research question will be answered: 'Can trend features extracted from EMG and PPG data be used to accurately predict deterioration in early septic patients at the emergency department?'

It is expected that continuous monitoring of the respiratory and cardiovascular systems will provide valuable insights into the patient's condition, enabling the early identification of deteriorating patients and the initiation of treatment at an earlier stage. This can lead to a reduction in in-hospital mortality, a shorter length of hospital stay, and decreased healthcare costs.

Research objectives

The aim of this research is to develop a model that is able to predict deterioration among patients with early sepsis at the emergency department (ED) based on EMG and PPG waveforms. In order to ensure the effectiveness of this model, it is important to use robust and reliable features as input. Therefore, the first step in this research is to investigate how to obtain reliable features. Next, it will be determined which of these features are most useful for predicting deterioration and can be incorporated into the models. Finally the most appropriate machine learning models will be selected and trained.

2 BACKGROUND

2.1 Clinical background

Sepsis is a potentially life-threatening condition that occurs when the body's immune responds to an infection in a way that leads to organ damage and failure. There is evidence that the development of sepsis involves not only the infectious agent and the immune response, but also changes in coagulation, immune suppression, and organ function. Sepsis is a common problem in older individuals, and it disproportionately affects people with cancer and underlying immune suppression. Despite significant progress in the understanding of the pathophysiology of sepsis, advances in hemodynamic monitoring and resuscitation measures, sepsis remains a leading cause of morbidity and mortality in critically ill patients. [10, 11].

Sepsis has been recognized for some time, but it was not until the late 20th century that it was formally defined clinically. This was due to the lack of effective antimicrobials and supportive care, which meant that patients with sepsis did not survive long enough to be studied. In February 2016, the European Society of Intensive Care Medicine and the Society of Critical Care Medicine (SCCM) published new consensus definitions of sepsis and related clinical criteria (Sepsis-3). According to these definitions, "sepsis is a life-threatening condition characterized by a dysregulated host response to infection that leads to organ dysfunction". Organ dysfunction is measured using the Sequential Organ Failure Assessment (SOFA) score, which is an objective measure of the number and severity of organ dysfunction in six organ systems (respiratory, coagulatory, hepatic, cardiovascular, renal and neurological). The SOFA score can be used to measure individual or aggregate organ dysfunction [12], which can be seen in figure 2.1. [1]

Organ system	SOFA score				
	0	1	2	3	4
Respiratory, PO_2/FiO_2, mmHg (kPa)	≥400 (53.3)	<400 (53.3)	<300 (40)	<200 (26.7) with respiratory support	<100 (13.3) with respiratory
Coagulation, Platelets, $\times 10^3/mm^3$	≥150	<150	<100	<50	<20
Liver, Bilirubin, mg/dL	<1.2	1.2–1.9	2.0–5.9	6.0–11.9	>12.0
Cardiovascular	MAP ≥70 mmHg	MAP <70 mmHg	Dopamine <5 or dobutamine (any dose)[b]	Dopamine 5.1–15 or epinephrine ≤0.1 or norepinephrine ≤0.1[b]	Dopamine >15 or epinephrine >0.1 or norepinephrine >0.1[b]
Central nervous system, Glasgow Coma Scale	15	13–14	10–12	6–9	<6
Renal, Creatinine, mg/dL. Urine output, mL/d	<1.2	1.2–1.9	2.0–3.4	3.5–4.9 <500	>5.0 <200

Figure 2.1: The Sequential Organ Failure Assessment (SOFA) score [11]

2.1.1 Pathophysiology of sepsis

It has been previously believed that the hemodynamic features of sepsis were largely a result of the hyperactive immune response of the host to a specific pathogen. However, molecular studies have revealed a far more nuanced and complex interplay between the infectious agent and host that together produce the manifestations of sepsis. [11]

Innate immunity and inflammatory mediators

In sepsis, there is an excessive immune response that leads to the cell death of host cells and tissues due to collateral damage. This immune response is initiated by the activation of innate immune cells, such as macrophages, monocytes, neutrophils and natural killer cells, through the binding of pathogen-associated molecular patterns (PAMPs) or damage-associated molecular patterns (DAMPs) to specific receptors. This activates intracellular signaling pathways that lead to the transcription and release of pro-inflammatory cytokines, including TNF-a, IL-1, and IL-6. Some pattern recognition receptors, such as NOD-like receptors, can also form inflammasomes that produce cytokines such as IL-1b and IL-18, as well as caspases involved in programmed cell death. Pro-inflammatory cytokines stimulate the activation and proliferation of leukocytes, activate the complement system, increase the expression of endothelial adhesion molecules and chemokines, induce the production of tissue factor and upregulate hepatic acute phase reactants. [11]

Dysregulation of hemostasis

In sepsis, there is an intersection between the inflammatory and hemostatic pathways, with the simultaneous activation of both the inflammatory and the coagulation cascades. This can range in severity from mild thrombocytopenia to disseminated intravascular coagulation (DIC). The hypercoagulability associated with sepsis is thought to be driven by the release of tissue factor from disrupted endothelial cells, which activates the coagulation cascade and leads to the production of thrombin, activation of platelets, and formation of platelet-fibrin clots. These microthrombi can cause local hypoperfusion, resulting in tissue hypoxia and organ dysfunction. In addition to the proagulant effects described above, sepsis also leads to a depression of the anticoagulant effects of protein C and antithrombin, which normally temper the coagulation cascade. Protein C is normally converted to activated protein C by thrombomodulin, which itself is activated by thrombin. Activated protein C has both anticoagulant effects through the degradation of factors Va and VIIIa in conjunction with activated protein S, and anti-inflammatory effects, including the inhibition of TNF-a, IL-1b and IL-6 and the limitation of neutrophil and monocyte adhesion to endothelium. In patients with sepsis, there are decreased levels of protein C, downregulation of thrombomodulin, and low levels of protein S, leading to the unregulated propagation of the coagulation cascade. [13]

Immunosuppresion

The initial proinflammatory state of sepsis is often superseded by a prolonged state of immunosuppresion. This is characterized by a decrease in the number of T cells due to apoptosis and a decreased response to inflammatory cytokines [14]. Postmortem studies of ICU patients who died of sepsis have shown a global depletion of CD4+ and CD8+ T cells, particularly in lymphoid organs such as the spleen [15]. Studies have also revealed a decreased production of crucial cytokines, such as the IL-6 and TNF-a in response to endotoxin [16]. In septic patients, neutrophils have been found to express fewer chemokine receptors and have diminished chemotaxis in response to IL-8 [17]. These findings suggest that the immune system in septic individuals is unable to mount an effective response to secondary bacterial, viral or fungal infections.

Cellular, tissue and organ dysfunction

The underlying mechanism behind tissue and organ dysfunction in sepsis is decreased oxygen delivery and utilization by cells due to hypoperfusion. Hypoperfusion is a result of cardiovascular dysfunction, which is common in sepsis. The incidence of septic cardiomyopathy ranges from 18% and 60%

in different studies and it thought to be related to circulating cytokines such as TNF-a and IL-1b, which can depress cardiac myocytes and interfere with their mitochondrial function. However, the impact of myocardial depression on mortality has not been established yet. In addition to cardiovascular dysfunction, sepsis also produces a state of hypotension and distributive shock due to arterial and venous dilation (induced by inflammatory mediators) and reduced venous return. This dilation affects all components of the microvasculature: arterioles, venules, and capillaries. This is exacerbated by the leakage of intravascular fluid into the interstitial space due to loss of endothelial barrier function, induced by alterations in endothelial cadherin and tight junctions. All of these changes in hemodynamics, in conjunction with microvascular thrombosis, can result in hypoperfusion in tissues and organs. This leads to increased anaerobic glycolysis in cells, resulting in the production of lactic acid and the production of reactive oxygen species (ROS). These ROS cause dysfunction of mitochondria and a drop in ATP levels. These mechanisms cause damage at the cellular levels. The broader alterations that occur in tissues and organs contribute to the morbidity and mortality of sepsis. [11]

There are significant alterations to the endothelium in sepsis, including disruption of its barrier function, vasodilation, increased leukocyte adhesion and the creation of a proagulant state. These changes lead to the accumulation of edema fluid in the interstitial spaces, body cavities, and subcutaneous tissue. In the lungs, there is a disruption of the alveolar-endothelial barrier and the accumulation of protein-rich fluid in the interstitial lung spaces and alveoli, which can cause a ventilation-perfusion mismatch, hypoxia, and a decreased lung compliance, potentially leading to the development of acute respiratory distress syndrome (ARDS) in severe cases. In the kidneys, a combination of reduced renal perfusion, acute tubular necrosis, and other defects in the microvasculature and tubules can result in varying degrees of acute kidney injury. In the gastrointestinal tract, the increased permeability of the mucosal lining leads to both bacterial translocation across the bowel and autodigestion of the bowel by luminal enzymes. In the liver, there is suppression of bilirubin clearance, leading to cholestatis. Altered mentation, a common symptom in sepsis, is indicative of central nervous system (CNS) dysfunction. The endothelial changes described above undermine the blood-brain barrier, allowing the entry of toxins, inflammatory cells, and cytokines and leading to cerebral edema, neurotransmitter disruption, oxidative stress, and white matter changes that can range from mild confusion to delirium and coma (septic encephalophathy). [11]

Treatment strategies

The success of treatment for sepsis depends on early diagnosis, prompt initiation of appropriate antibiotic treatment and elimination or recovery from the underlying disease. Antibiotics are often administered empirically while awaiting determination of the microorganism. The first six hours after the onset of sepsis symptoms are particularly critical for prognosis, and it has been shown that appropriate antibiotic treatment within this time frame can reduce the incidence of shock by half, regardless of the underlying disease in cases of sepsis caused by gram-negative bacteria. [18]

The primary goal in septic shock is to regulate blood volume and provide sufficient tissue perfusion. The first step in achieving this is to administer appropriate fluid therapy. In patients who have achieved fluid balance, but still have hypotension, vasoactive drugs may be added to the treatment. Corticosterioids are another important agent that have been shown to reduce the risk of death in the treatment of sepsis. [18]

2.2 Technical background

In this study, photoplethysmography (PPG) and diaphragm electromyography (EMG) were utilized for data acquisition. The principles and applications of these techniques will be described in the following sections. Additionally, trend analysis will be analyzed and a range of machine learning models will be explored.

2.2.1 Photoplethysmography

PPG is a non-invasive technique that uses light to measure changes in the volume of blood in a specific region of the body, typically the finger or earlobe, as depicted in Figure 2.2a and 2.2b. A light source is used to illuminate the tissue, and a photodetector measures the variations in light intensity resulting from changes in perfusion during the cardiac cycle, as illustrated in Figure 2.2c. The wavelength of the light used in PPG plays a significant role in light-tissue interaction. The primary constituent of tissue, water, strongly absorbs light in the ultraviolet and longer infrared wavelengths. Shorter wavelengths of light are also absorbed by melanin to a considerable extent. However, there is a window in the absorption spectrum of water that allows visible (red) and near infrared light to pass through more easily, enabling the measurement of blood flow or volume at these wavelengths. As a result, infrared (IR) or near infrared (NIR) light is often chosen for use as the PPG light source. [19]

The photoplethysmography (PPG) waveform consists of both an alternating current (AC) component and a direct current (DC) component, as depicted in Figure 2.2c. The AC component is the pulsatile component of the PPG waveform and reflects the changes in blood volume that occur during the cardiac cycle. It typically has a fundamental frequency of around 1 Hz, depending on the heart rate. The DC component reflects the average blood volume and the tissue, and changes slowly due to factors such as respiration, vasomotor activity, and vasoconstrictor waves. The PPG waveform is characterized by two phases: the anacrotic phase, which corresponds to the rising edge of the pulse, and the catacrotic phase, which corresponds to the falling edge of the pulse. The first phase is primarily associated with systole, while the second phase is associated with diastole and wave reflections from the periphery. A dicrotic notch is typically observed in the catacrotic phase of individuals with healthy, compliant arteries. [19, 20]

In a pulse oximeter, the ratio of NIR and IR light absorbed by the blood is typically measured using a pair of LED light sources, one emitting NIR light and the other emitting IR light. The intensity of light that reaches the detector is measured for each wavelength, and the ratio of the two intensities is used to calculate the oxygen saturation of the blood. However, in this research, the focus is not on measuring oxygen saturation, but on the analysis of the PPG waveform itself. [21]

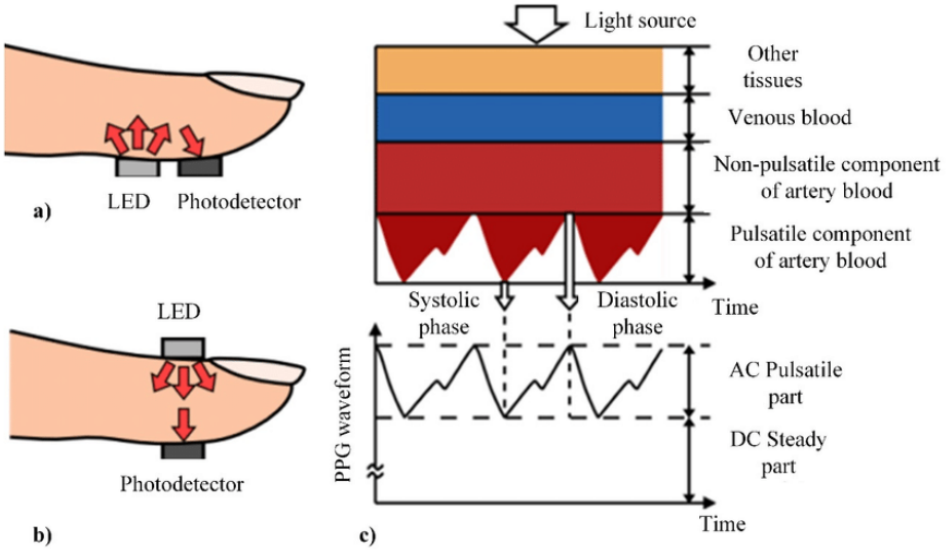


Figure 2.2: Principle of photoplethysmography [22]

PPG features

The derived PPG features are divided into five categories: amplitude dependent features, time dependent features, areas, indexes and ratios, and second derivative features, also known as acceleration

plethysmography. Most of the features are visualized in Figure 2.4.

Amplitude dependent features
There are three amplitude dependent features. One commonly used feature is the Systolic Peak Amplitude (SPA), which is defined as the magnitude of the first peak within a single PPG waveform. The SPA serves as a quantitative indicator of the pulsatile changes in blood volume that result from arterial blood flow around the measurement site. SPA has been linked to be correlated with stroke volume [23]. Furthermore, it has been proposed that the SPA may be an appropriate measure than pulse arrival time for estimating continuous blood pressure [24]. The second amplitude-dependent feature is the Diastolic Peak Amplitude (DPA), which corresponds to the magnitude of the second peak of the PPG waveform. The DPA is considered as the late systolic peak, as it primarily arises from the reflected systolic wave from the lower body [20]. Lastly, the Relative Amplitude (RA) is calculated by dividing the SPA by the true value of the systolic peak.

Time dependent features
In this study, six time-dependent features were extracted from the PPG signals. The Pulse Width (PW) is calculated at the point of half-height of the systolic peak. There is suggested that the PW is more strongly correlated with the systemic vascular resistance than the SPA [25]. The Peak-to-Peak (PtP) interval represents the time between two consecutive systolic peaks and is highly correlated with heart rate [26]. The Pulse Interval (PI) is defined as the time between the onset and the end of a PPG wave. The Crest Time (CT) corresponds to the time from the onset of the PPG waveform to the systolic peak amplitude and it has been demonstrated that the CT can be a useful feature for cardiovascular disease classification [27]. The Delta Time (DT) is defined as the time between the systolic peak and diastolic peak, and it relates with the time it takes for the systolic pressure to propagate from the heart to the periphery and back [28]. Lastly, the Heart Rate (HR) is based on the pulse interval.

Areas
Stroke volume (SV) can be indirectly measured by calculating the area under the curve (AUC) of the systolic segment of the PPG waveform, as illustrated by area A1 in Figure 2.4. Although this feature has not been used in previous studies, a similar calculation method is used for arterial blood pressure (ABP) waveforms [29]. Given the similarity between ABP and PPG waveforms, the SV feature has been incorporated into this study as well. Subsequently, cardiac output (CO) can be computed by multiplying the calculated SV with the heart rate.

Indexes and ratios
The ratio between the DPA and the SPA is called the reflection index (RI), because it is the ratio between the original wave and the reflected wave [20]. The augmentation index (AI) is the normalized version of the RI, which is specified as the SPA minus the DPA, divided by the SPA [20]. The area under the curve (AUC) can be divided into two parts at the dicrotic notch, as visualized in Figure 2.4. The infliction point area (IPA) ratio is the ratio between two areas and is specified as the area under the diastolic peak (A2) divided by the area under the systolic peak (A1) [28]. Wang et al., found that the ratio of these two areas can be used as an indicator of total peripheral resistance [30].

Acceleration Plethysmography
The second derivative of photoplethysmogram, also known as the acceleration plethysmogram (APG), is an indicator of the acceleration of blood in the finger. The APG waveform includes four systolic waves (a,b,c,d) and one diastolic wave (e-wave), which can be seen in Figure 2.3.The e-wave represents the dicrotic notch and corresponds to the closure of the aortic valve and subsequent retrograde blood flow. The ratio between these waves can be determined. However, not every subject has all the different waves, so a uniform feature for all subjects, the b/a ratio, was chosen. Takazawa et al. [31] demonstrated that the b/a ratio reflects increased arterial stiffness, hence the b/a ratio increases with

age. Imanaga et al. [32] provided a direct evidence that that the magnitude of b/a of the APG is related to the distensibility of the peripheral artery and showed that it is a useful index of atherosclerosis and altered arterial distensibility.

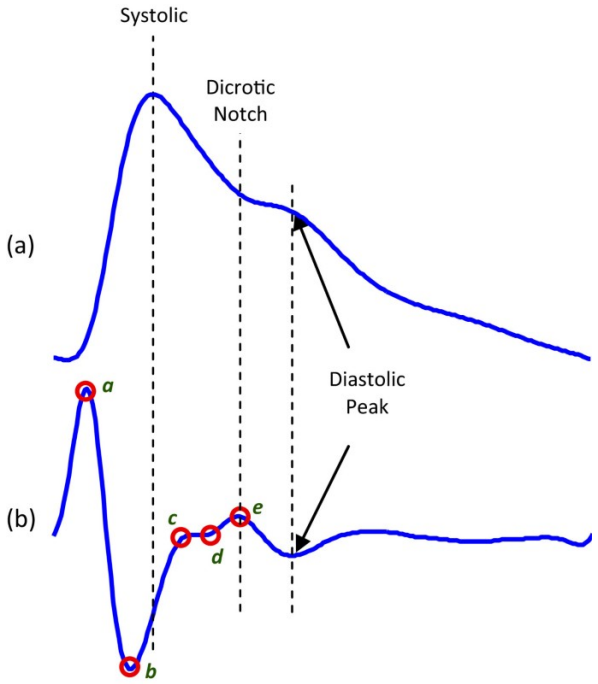


Figure 2.3: The PPG waveform is presented in part (a) of the figure, while part (b) illustrates the second derivative of the PPG waveform. Part (a) allows for the identification of the a, b, c, d, and e waves. [33]

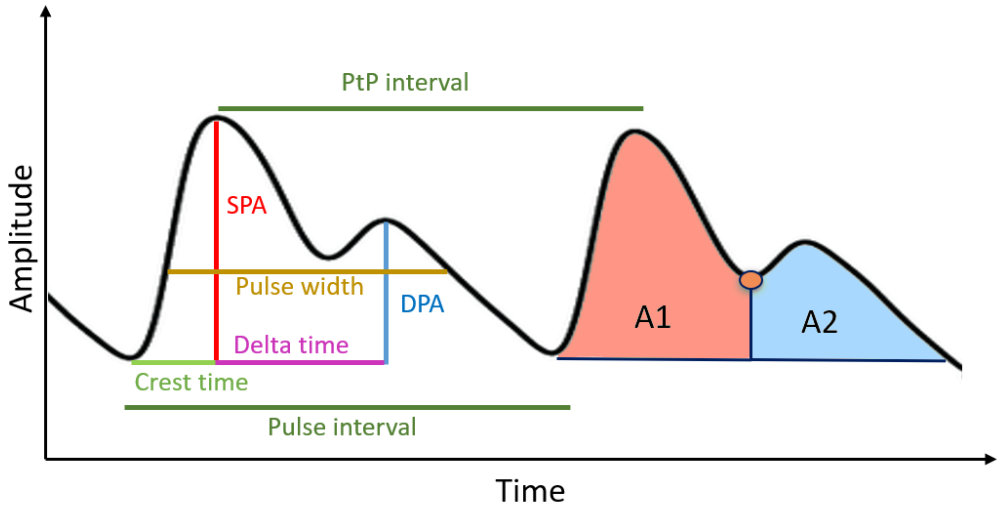


Figure 2.4: Overview of PPG features. SPA is the systolic peak amplitude, DPA is the diastolic peak amplitude, A1 is the area under the systolic part of the curve, A2 is the area under the diastolic part of the curve.

2.2.2 Electromyography

Electromyography (EMG) is a technique that records the electrical potentials produced during muscle contractions. It uses electrodes placed on the skin or inserted into muscle tissue. Surface EMG is the most commonly used method, as it is non-invasive and more practical to use than intramuscular EMG.

Respiratory activity can be monitored using surface EMG of the main respiratory muscle, the diaphragm. The diaphragm separates the thoracic cavity and the abdominal cavity. Upon inhalation, the diaphragm contract and flattens, causing the chest cavity to expand. This contraction creates a vacuum, which draws air into the lungs. Upon exhalation, the diaphragm relaxes and returns to its domelike shape, forcing air out of the lungs [34]. Surface EMG recordings of the diaphragm pose some challenges as the diaphragm is not a subcutaneous muscle. Signal power loss occurs through the tissues interposed between the electrodes and the diaphragm. However, it has been determined that quality EMG recordings, minimally affected by unwanted factors, can be obtained from the diaphragm with appropriate electrode placement [35]. An increase in respiratory rate will be reflected as more frequent activation of the diaphragm muscle and it is hypothesized that an increase in the work of breathing will likely result in an increased area under the curve [36].

2.2.3 Trend analyses

Trend analysis is a technique used to identify patterns and trends in intermittent recorded vital signs or waveforms by examining changes over time. This approach is particularly useful in detecting early signs of deterioration in a patient's condition, which can be missed by relying on individual measurements. Despite their potential usefulness, trend analysis are not widely used due to a lack of knowledge and appreciation among doctors [9]. However, a systematic review by Brekke et al., found that trend analysis of intermittent vital signs can increase the accuracy of detecting clinical deterioration in both the ED and on general wards [9, 8]. Various trend features, including the change in the current value from the previous value, the mean, standard deviation, slope, minimum and maximum, were investigated.

In a other study by Bloch et al., machine learning models were trained using a set of features representing the variability in continous measured vital signs at the ICU, with the hypothesis that changes in the variability of certain physiological parameters could serve as an early warning for impending sepsis. One of the trend features used in the study was the number of trend changes, defined as a series of local minimum and maximum values. A vital sign with a higher number of trend changes is considered less stable than one with fewer changes. Other trend features included the mean, minimum, and maximum intensity of changes, which indicate the magnitude of changes in a vital sign. A vital sign with a higher intensity of change is considered less stable than one with a lower intensity. These features took into account both the amount and intensity of changes in vital signs. The study found that variability in vital signs was a good indicator of the risk of sepsis during an ICU stay [37].

The results of previous studies have shown that adding trends generally lead to increased accuracy in detecting clinical deterioration compared to models relying solely on current vital signs. However, the generalililizability of these findings is difficult to assess and there is no consensus on the most effective methods for trend analysis or the optimal features to use as input for predictive models.

2.2.4 Machine-learning

In recent years, machine learning has gained widespread use in the medical field due to its ability to process large amounts of data and provide accurate predictions of patient outcomes. Machine learning is a subfield of artificial intelligence (AI) that focuses on developing systems that learn patterns from data in a way that is not explicitly programmed. These systems improve their performance on a specific task through iterative computations that enable them to adapt to new situations. Machine

learning can be divided into three categories: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning uses a labeled dataset, which is typically split into a training and test set for model training and validation. The resulting supervised learning model is then used to generate predictions on previously unseen, unlabeled data. Examples of supervised learning include classification and regression. Unsupervised learning involves identifying hidden patterns and relationships in an unlabeled dataset through grouping data into clusters or through association. Finally, reinforcement learning is a type of machine learning in which the learner is an agent in an environment that learns through taking actions and receiving feedback on the effectiveness of these actions in solving a particular problem. [38]

Logistic regression

Logistic regression is a supervised machine learning algorithm used for binary classification problems. While linear regression can predict a continuous variable based on the weight of each independent variable, logistic regression is used for classifying data into two distinct categories. It follows the same principle as linear regression, where the prediction is determined by the summation of the product of features and their corresponding weight:

$$y = B_0 + B_1 x_1 + B_2 x_2 \tag{2.1}$$

where the dependent variable y depends on the sum of the intercept B_0 and for each independent variable, the multiplication of the variable's value x and its weight B. It then uses a logistic function that computes the combination of the independent variables, resulting in the logarithmic odds for a sample belonging to a certain class. This logistic model can be seen in Equation 2.2.

$$p = \frac{1}{1 + e^{-(B_0 + B_1 x_1 + B_2 x_2)}} \tag{2.2}$$

where given an input sample x, it computes the linear equation 2.1 and returns the probability of this input belonging to a class. The parameters of the intercept and weight are optimized during the training process so that the model can correctly classify new examples [39].

K-Nearest Neighbours

The K-Nearest Neighbors (KNN) algorithm uses distance calculations to identify instances with similar proximity. When developing a KNN model, the first step is to determine the number of nearest neighbours, also known as K. A small K value results in a higher influence on the output, while a higher K can lead to a smoother decision boundary with lower variance, but increased bias. One approach to selecting K is to train a model with various K values and use a cross-validation technique to determine which K results in the highest testing accuracy. Next, a distance function such as Euclidean distance is used to calculate the distance between each instance and all training samples. The K neighbors with the least distance are chosen, and the category or numeric value of these nearest neighbors is obtained. Finally, the class of an instance is predicted based on the majority of the nearest neighbors. [39]

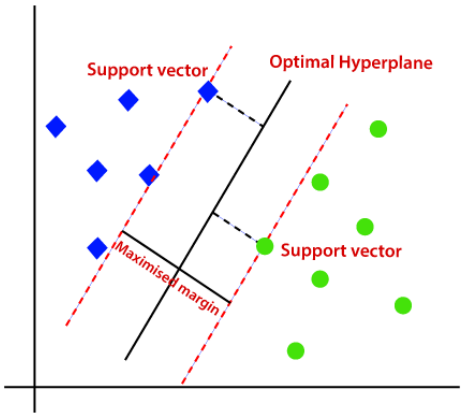


Figure 2.5: Support Vector Machine for a two dimensional space [40]

Support vector machine

Support Vector Machines (SVM) are another form of supervised machine learning algorithms that can be used for both classification and regression tasks. The goal of SVM is to find a hyperplane that maximally separates different classes in the data by maximizing the margin between the hyperplane and the nearest data points, known as support vectors. This is illustrated in Figure 2.5, which shows

a separating hyperplane (black solid line) drawn to distinguish between blue and green instances in a two-dimensional space. In a three-dimensional space, the hyperplane becomes a two-dimensional plane. The distance between the dashed red lines and the separating hyperplane is referred to as the margin. The objective in SVM is to maximize this margin from both sides. [39]

Decision Tree and Random Forest

A decision tree is a machine learning algorithm that utilizes a set of features in the data to make a series of decisions and reach a specific result. The order in which these features are evaluated is based on criteria such as the Gini impurity index or information gain [39]. An ideal decision tree would accurately predict class labels. However, decision trees have the potential to overfit, meaning they may perform well on the training data, but poorly on new data. To reduce this risk of overfitting, the Random Forest classifier can be used. Random Forest is an ensemble of decision trees, as shown in Figure 2.6. While a single decision tree may produce inaccurate predictions due to variance, taking the majority vote from hundreds or thousands of trees in a Random Forest can reduce this variance and improve the accuracy of the final prediction. [39]

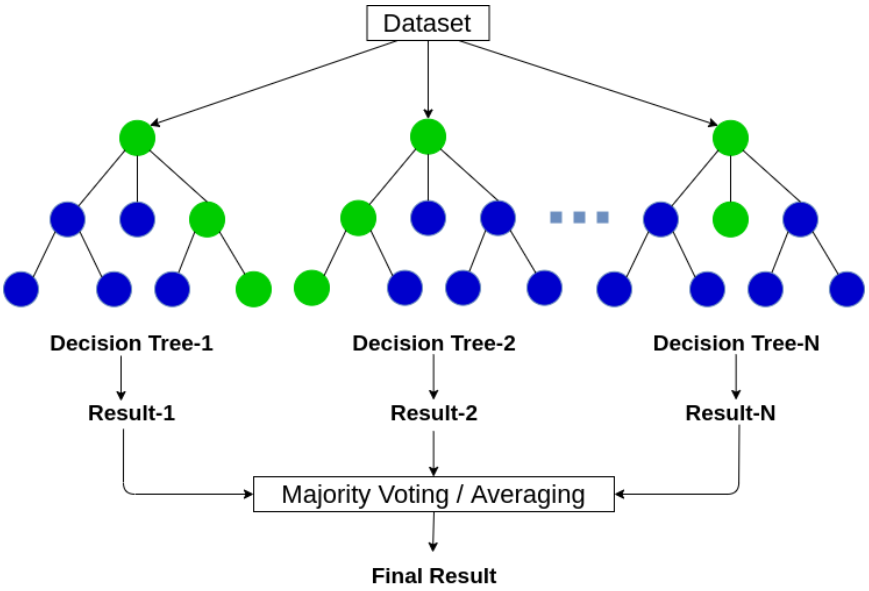


Figure 2.6: Random forest [41]

Feature selection

There are several strategies that can be used to select relevant features as input for machine learning models. One approach is to remove highly correlated features, as these may not provide additional value to the model and may even be redundant. Alternatively, wrapper methods, which involve the use of a machine learning algorithm to identify the most relevant features, can be used. These methods iteratively evaluate the performance of the model using different subsets of features in order to determine the optimal subset. Filter methods, on the other hand, involve the selection of features based on statistical tests and are generally independent of the model.

Data augmentation

To address imbalance in a dataset, Synthetic Minority Oversampling (SMOTE) can be used. SMOTE generates new minority samples by selecting a random minority class instance and determining its k-nearest neighbors. A random instance is then chosen from these neighbors and a synthetic example is created at a randomly determined point between the two samples in the feature space. This process is repeated until there are as many minority samples as there are of the original majority class.

Cross-validation

Cross-validation is a technique used in machine learning to assess the performance of a model on an

independent data set. In K-fold cross-validation the data set is divided into k equally sized folds. The model is trained on k-1 of the folds and the remaining fold is used as the test split. This procedure is then repeated k times, with a different fold being used as the test set each time. The average performance measure across all k iterations is used as the performance measure of the model.

Performance measures

There are several performance measures that can be used to evaluate the performance of classifiers, including accuracy, precision, recall, AUC, and F1 score. These measures are all derived from the confusion matrix, which can be seen in Figure 2.7. In this study, the goal is to predict as many data points correctly as possible, which is measured by accuracy. However, accuracy may not provide enough information, especially for imbalanced classes. For example, if a dataset has 100 samples with 90 samples belonging to class 0 and 10 samples belonging to class 1, and the model predicts that every sample belongs to class 0, the accuracy would be 90%, but the model would not be useful as a classifier. Therefore, precision and recall are also considered. Precision measures the proportion of positive predictions that are correct, while recall measures the proportion of actual positive cases that were correctly predicted. A model with a precision of 100% means that no subjects were falsely marked as a risk for deterioration, while a model with a recall of 100% means that all subjects that will deteriorate are correctly labeled.

It is rare for a model to have perfect scores for both specificity and recall (sensitivity), so there is often a trade-off between these two measures. This balance can be visualized using a Receiver Operating Characteristic (ROC) curve, which plots the true positive rate (sensitivity) against the false positive rate (1-specificity). The AUC value, or the area under the ROC curve, can be used as a summary statistic for comparing classifiers. A perfect classifier has an AUC of 1.0, while a random classifier has an AUC of 0.5. However, AUC may not be suitable for evaluating classifiers on imbalanced data sets, as it is not sensitive to the relative costs of false positives and false negatives. In such cases, the F1 score may be a better measure. The F1 score is the harmonic mean of precision and recall, and it takes into account both false positives and false negatives as can be seen in equation 2.3. This makes it useful for evaluating classifiers on imbalanced data sets.

$$F1_{score} = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{2.3}$$

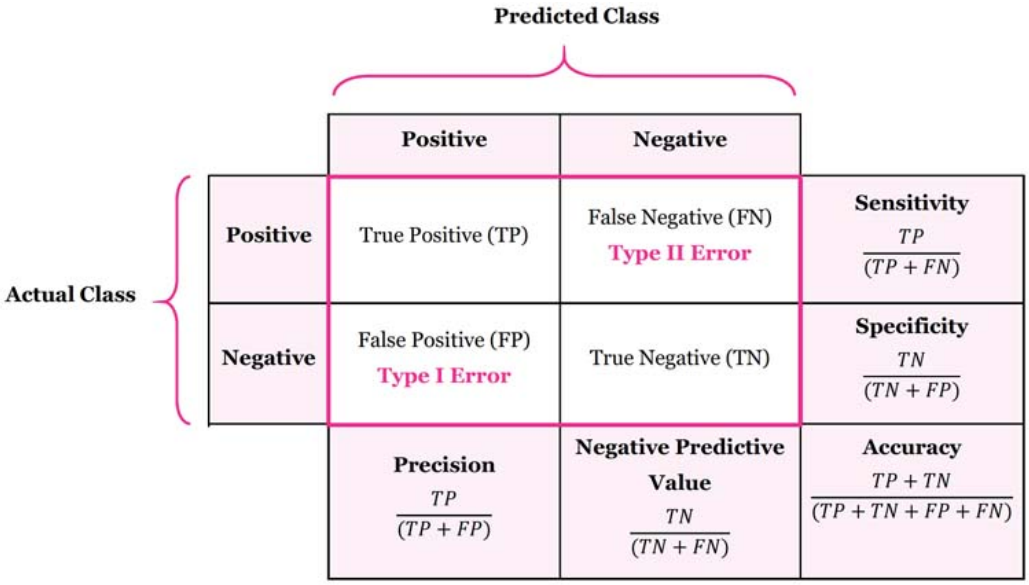


Figure 2.7: Confusion matrix [42]

3 METHODS

This chapter presents the methodology of the study, which is divided into six parts. The first part provides an overview of the dataset and the inclusion criteria. The second and third sections present and discuss the preprocessing of the PPG and EMG waveforms, respectively. The fourth section explains the trend analysis process. The fifth part elaborates on the statistical analysis and the sixth section outlines the methodology for the machine learning models.

3.1 Dataset

This study uses vital signs and continuously measured PPG and EMG data collected by the Acutelines data and biobank project at the ED of the University Medical Center Groningen (UMCG) [43]. The Acutelines project aims to gather data and biomaterials from critically ill patients in order to facilitate research to the pathophysiology, diagnosis, treatment and prognosis of severe diseases. The bank includes bedside monitoring data, such as electrophysiological waveforms and vital parameters, as well as information from other sources such as electronic health records.

3.1.1 Population

Patients were included who visited the ED of the UMCG between 09.00 and 23.00 from March 2021 to October 2022 who met the following inclusion criteria:

- Inclusion criteria:
  - Adults aged 18 years or older
  - Consent to use their data for research
  - Meeting one of the following (early) sepsis criteria of Acutelines:
    - Clinical suspicion of sepsis by a physician or nurse
    - Two or more SIRS criteria (Figure 3.1)
    - Two or more qSOFA criteria (Figure 3.2)
    - Blood cultures in combination with Emergency Severity Index (ESI) [44] level 1 or 2 or blood cultures taken in combination with admission by ambulance and ESI level 3.
    - Temperature $< 35\ ^{\circ}$C or $> 38\ ^{\circ}$C in combination with ESI level 1 or 2 or temperature $< 35\ ^{\circ}$C or $> 38\ ^{\circ}$C in combination with admission by ambulance and ESI level 3.
- Exclusion criteria:
  - Known cardiac arrhythmia's
  - Implanted pacemakers, deep brain stimulators or other stimulation devices implanted under the skin in the thoracic or abdominal area
  - Comorbidities with known signal disturbance, such as pneumothorax and ascites
  - No admission to the nursing ward at the UMCG

Systemic Inflammatory Response Syndrome
Temperature >38.3°C, or <36°C
Heart Rate >90 bmp
Respiratory rate >20 breaths/min
White cell count <4 or >12 g/L
Blood glucose >7.7 mmol/L not diabetic
New altered mental state

Figure 3.1: SIRS criteria [45]

qSOFA (Quick SOFA) Criteria	Points
Respiratory rate ≥22/min	1
Change in mental status	1
Systolic blood pressure ≤100 mmHg	1

Figure 3.2: qSOFA criteria [46]

The patients were divided into two groups: deteriorated and not deteriorated. A deterioration label was assigned to a patient if they fell into one of the following categories within a 48-hour window after ED admission:

- Deceased
- ICU admission
- Increase of two points in SOFA score in combination with a clinical proven infection based on the performed adjudication

The SOFA score was calculated according to the criteria outlined in Figure 2.1. Many of these criteria are based on laboratory test results, but not all tests were performed on all patients. In cases where a value was not available, it was assumed that there was no clinical relevance for this measurement and the normal value was used in calculating the SOFA score.

The SOFA score was calculated for each patient at three time points: at admission (day 0), after 24 hours (day 1), and after 48 hours (day 2). When multiple laboratory values were available for calculating the SOFA score for a specific organ system, the value representing the most severe organ dysfunction was chosen. In accordance with the guidelines described in [1], only the difference in scores between time points was analyzed and the absolute SOFA score at each time point was not considered. The score differences between day 0 and day 1, day 1 and day 2 and day 0 and day 2 were calculated, and patients with an increase of two or more points on any of these comparisons were labeled as deteriorating.

3.1.2 Data acquisition

All patients who were eligible and consented to the wearables, were connected to the OSAsense (DiagnOSAS B.V., Haaksbergen, The Netherlands) and Sencure (Roden, the Netherlands) wearables, which measured the PPG and EMG signals. The PPG wearable consisted of a wristband with a monitor and a finger pulse oximeter, which can be seen in Figure 3.3. The display of the wristband shows the saturation, heart rate and PPG waveform. Besides these parameters, the OSAsense also determines the acceleration of the wristband. PPG signals are measured using red (660 nm) and infrared (910 nm) light. For PPG signal analysis, the infrared light signal was used, rather than the red light signal, as the amplitude of the infrared signal was higher due to its deeper penetration, caused by its longer wavelength [47].

The diaphragm EMG wearable consisted of three electrodes, with two electrodes placed right and left midclavicular on the costal margin and the third, reference electrode placed midsternal at the third intercostal space. This can be seen in Figure 3.4. The protocol for applying these wearables is added in the Appendix A.2.

The measurement started within six hours after admission to the ED and continued during the admission to the general ward or ICU and lasted about twelve hours. An example of the timeline can be seen in Figure 3.5. In case the patient only consented to one of the two wearables, only one wearable was connected. The protocols of the wearables can be seen in Appendix A.2.

Figure 3.3: OSAsense device [48]

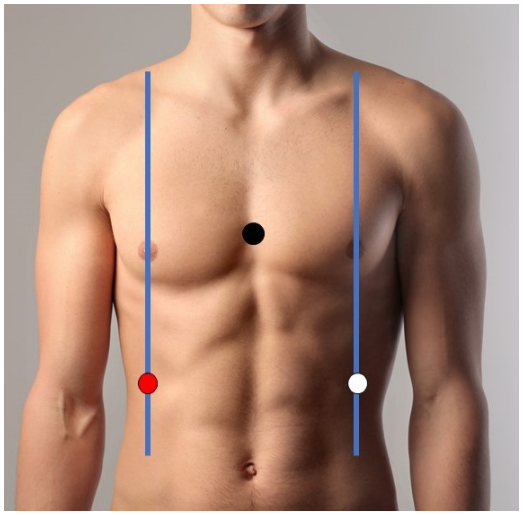


Figure 3.4: Location of diaphragm EMG electrodes.

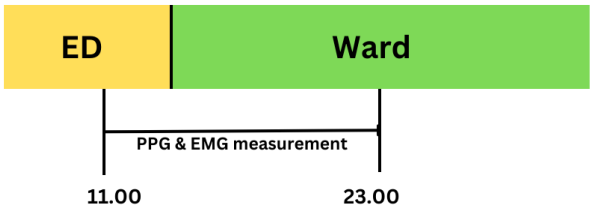


Figure 3.5: Example of timeline of EMG and PPG measurement

3.2 Photoplethysmography

This section begins by explaining the pre-processing steps, followed by a discussion of the chosen features. Lastly, the outlier removal process will be described

3.2.1 Pre-processing

The PPG pre-processing consists of six steps, shown in Figure 3.6, including basic pre-processing (interpolating, filtering), a first quality check, detection of the points of interest, a second signal quality check (SQI), feature calculation and outlier removal.

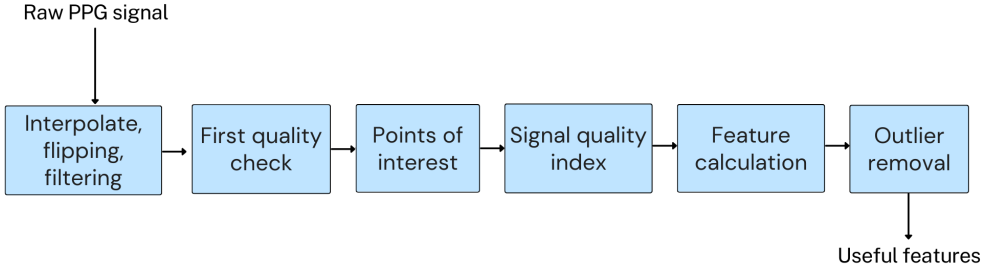


Figure 3.6: Flowchart of the PPG pre-processing process.

The sampling frequency of the OSAsense sometimes slightly deviates from the preset sample frequency. For this reason, the PPG data was first linearly interpolated to a fixed sample frequency of 50 Hz. The data was then flipped by subtracting it from the maximum value of the measurements ($1*10^6$), which transformed the signal to an absorbed PPG signal. This transformation resulted in a signal that adheres to the standard format of PPG signals, which is similar to the signals obtained from Philips monitors in the ED. In this way, the signals can be consistently preprocessed and compared. The PPG signal was then filtered using a 4th order Butterworth band-pass filter, with a frequency range of 0.8-20 Hz.

After filtering, a first quality check was executed to remove any data contaminated by movement artifacts. This was achieved by analyzing the movement of the wrist on a per-minute basis and identifying any minutes during which the deviation from the median gravitational force exceeded 0.02 g. These minutes were subsequently excluded from the analysis.

Prior to identifying points of interest, extreme artifacts were eliminated from the PPG signal through the use of a threshold based on the squared PPG signal and two moving averages of distinct durations. Specifically, two moving averages were computed, one with a duration of 5 seconds and the other with a duration of 60 seconds, both derived from the squared PPG signal. A threshold value was subsequently established by adding a constant to the 60 second-moving average. This threshold was then used to identify and remove any data points from the 5-second moving average that exceeded this threshold value.

Then, for each 20-second window, all peaks were selected, which included the systolic and diastolic peaks. To detect solely the systolic peaks, only those with amplitudes higher than the standard deviation of the peaks were saved. Onsets were determined by finding the depression before the systolic peaks. The location of the diastolic peak was defined as the first depression of the second derivative between the systolic peak and the following onset. For the dicrotic notch, the locations where the second derivative between the systolic peak and diastolic peak crosses zero was used. [20] The point of interest detection is visualized in Figure 3.7.

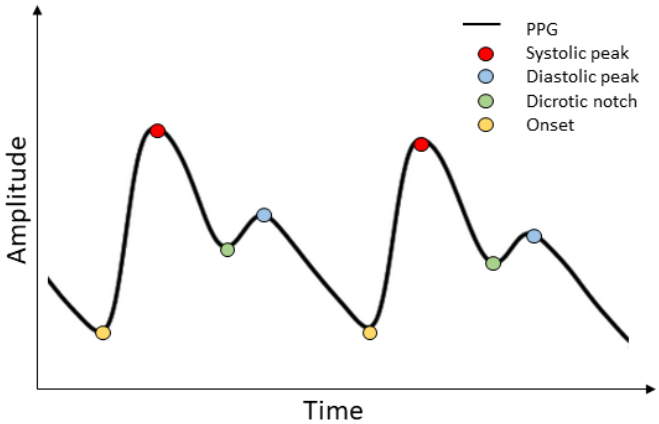


Figure 3.7: Point of interest detection of PPG signals where the red circles indicate the systolic peaks, the blue ones the diastolic peaks, the green ones the dicrotic notches and the yellow ones the onsets.

After all points of interest were detected, a more thorough signal quality check is performed using the Signal Quality Index (SQI) method adapted from Li and Clifford (2012) [49]. This method uses different templates and checks the correlation of each pulse with the templates to classify the pulse as good or bad. The first template is created by averaging all pulses within one minute and the length of the template was set to the average length of the pulses as can be seen in Figure 3.8. Pulses with correlations lower than 80% according to the first template are discarded. The second template is the average of the remaining pulses, see Figure 3.9. The correlation of each pulse with the mean of

the second template, Figure 3.10, is determined in three different ways. The first method uses the length of the template as a fixed length for the pulse. The second method resamples the pulse, so that the full pulse is matched with the template. In the third method, the pulse is resampled using the dynamic time warping method and the correlation is determined using this resampled pulse. A pulse is considered good if all three correlations are higher than 80% or if two of the three correlations had a correlation higher than 90%. If at least 50% of the pulses in one minute are good, the minute was considered a good and useful. An illustration of this can be observed in Figure 3.11, where the pulses deemed to be 'good' have a value of 5000, while remaining pulses have values of 2500 or 0.

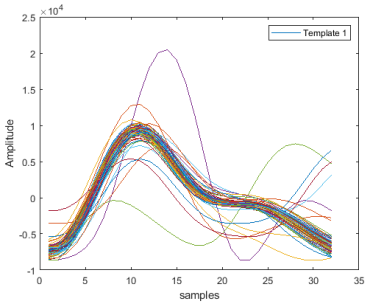


Figure 3.8: Template 1: All PPG waves in one minute.

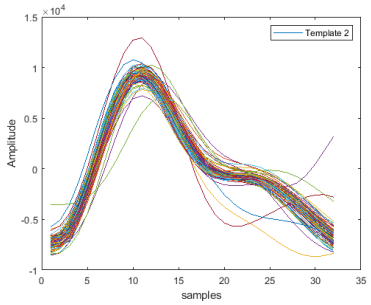


Figure 3.9: Template 2: Remaining waves after excluding waves.

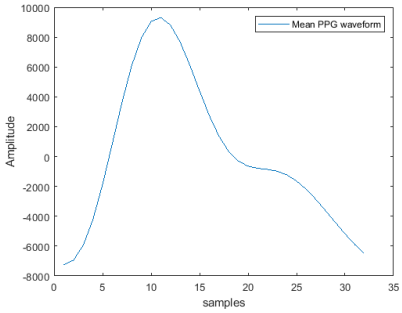


Figure 3.10: Mean PPG waveform

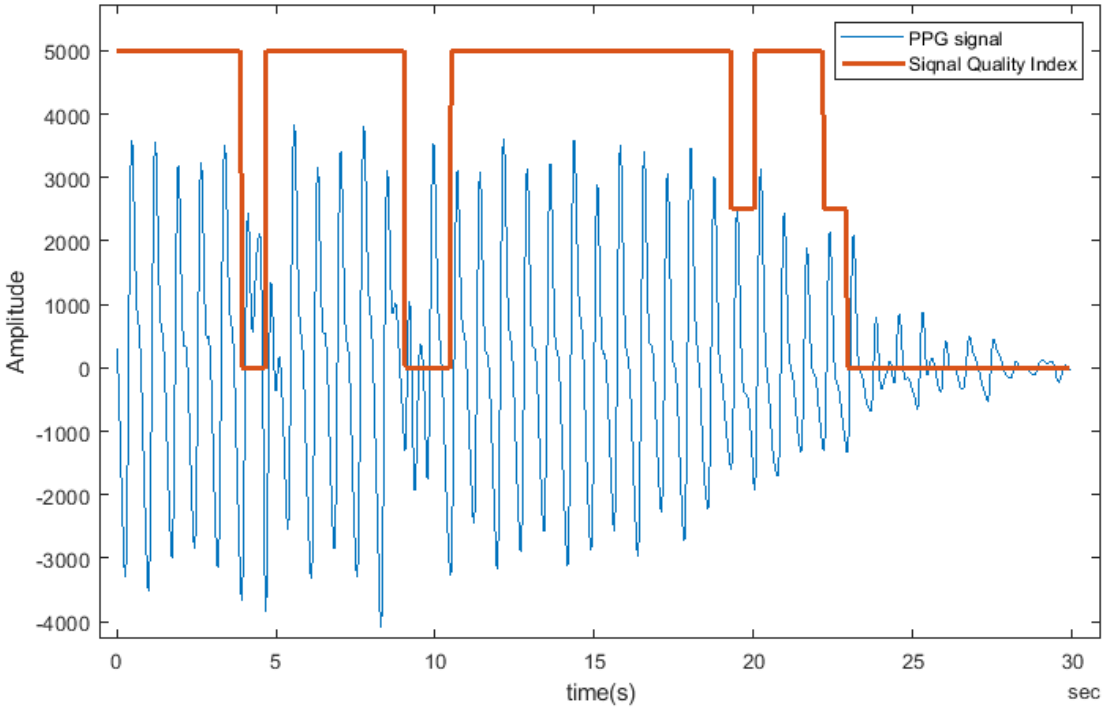


Figure 3.11: Signal Quality Index (red) based on template correlation. The waves with a SQI of 0 and 2500 are excluded from analysis.

Feature selection

Based on the determined points of interest, fifteen features are calculated. The features are described in the following Table 3.1 and can be seen in Figure 2.4 in the Background section.

Table 3.1: PPG features

Type Feature	Feature	Description
Amplitude	Systolic peak amplitude (SPA)	The amplitude after subtracting the onset height from the height of the systolic peak
	Diastolic peak amplitude (DPA)	The amplitude after subtracting the onset height from the height of the diastolic peak.
	Relative amplitude (RA)	The relative amplitude is determined by dividing the SPA by the true value of the systolic peak.
Time	Pulse Width (PW)	The pulse width is the width at the half height of the systolic peak.
	Peak-to-peak time (PtP time)	The time period between two consecutive systolic peaks.
	Pulse Interval (PI)	The time period between two consecutive onsets.
	Crest Time (CT)	The time period from the onset of the PPG waveform to the systolic peak
	Delta Time (DT)	The time period between the systolic peak and diastolic peak.
	Heart Rate (HR)	The heart rate is calculated by dividing 60 by the pulse interval.
Areas	Stroke Volume (SV)	An indirect measure of stroke volume is derived by calculating the AUC of the systolic part of the PPG waveform. (Area A1 in Figure 2.4)
	Cardiac Output (CO)	Cardiac output is calculated by multiplying the stroke volume with the heart rate.
Indexes and ratios	Reflection Index (RI)	The ratio between the DPA and the SPA.
	Augmentation Index (AI)	Normalized version of the RI, specified as the SPA minus the DPA, divided by the SPA.
	Infliction Point Area (IPA)	The area under the curve (AUC) is divided into two parts at the dicrotic notch, as visualized in Figure 2.4. The IPA is the ratio between these two areas and is specified by area under the diastolic peak (A2) divided by the area under the systolic peak (A1).
Acceleration Plethysmography	b/a ratio	The ratio between the b-wave and the a-wave of the second derivative of the PPG waveform.

Feature normalization

In order to account for differences in the size and shape of the finger and differences in the distance between the sensor and skin, various amplitude related features of the PPG signal are normalized. Normalizing the PPG signals allows for a more accurate comparison of the relative changes in the PPG signal over time between patients, rather than being influenced by differences in the absolute values of the PPG signals. The systolic, diastolic peak amplitude and stroke volume depend directly on the amplitude of the signal and are therefore normalized in the following way:

$$x_{normalized} = \frac{x_n - x_{min}}{x_{max} - x_{min}} \tag{3.1}$$

where x_{min} is the average of the lowest 25% in the dataset and x_{max} is the average of the highest 25% in the dataset. To minimize the influence of any potential artifacts in the data, it was decided not to use the absolute highest and lowest values as x_{min} and x_{max}.

Outlier removal
Only features of good quality waveforms are used for analysis. Despite this, some outliers may still remain in the data. To remove these outliers, the data is examined per feature and per hour. Feature values higher or lower than the median plus ten times the standard deviation of a feature are defined as outliers and are removed from the analysis. This process may result in some pulses where only a few features are calculated. These pulses are likely to be unreliable, so all features from a pulse are removed if more than six features are absent.

3.3 Diaphragm EMG

The pre-processing of the EMG signal is divided into two parts due to its complexity and the extensive numbers of steps involved. The first part involves a series of steps to remove cardiac activity from the EMG signal, including filtering, a first cardiac artifact removal, calculating the Root Mean Square (RMS), a second cardiac artifact removal and resampling. These steps are illustrated in Figure 3.12. The second part of the pre-processing focuses on removing remaining artifacts and segments of poor quality from the ECG filtered signal in order to ensure reliable feature determination. Following these pre-processing steps, the chosen features will be discussed.

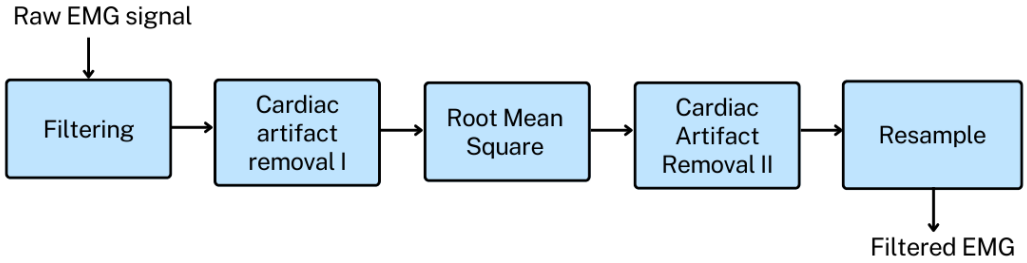


Figure 3.12: Flowchart of the EMG pre-processing part 1

3.3.1 Pre-processing part 1

To retrieve the respiratory EMG signal, the onset, drift and high-frequency noise are first removed using a second-order Butterworth band-pass filter with a frequency range of 20-150 Hz. However, the significant difference in strength between the electrical activity of the heart and respiratory muscles can make it difficult to identify the respiratory EMG signal. To address this, a method is used to reduce the impact of cardiac activity while preserving the signal of respiratory muscle activity. In this study, the template subtraction technique, as described by Petersen et al, was chosen [50, 51]. The R-peaks are first detected using the Pan-Tompkins method [52]. A template of the QRS complex is then created by averaging the QRS complexes (a window around the R-peak). The template's offset was made adaptive, and a least squares algorithm was used to adjust its height and width during the data analysis. Once the template was properly scaled, it was aligned with the detected peaks and subtracted from the EMG signal. The removal of the cardiac activity can be seen in the top two figures or Figure 3.14. The process of removing cardiac activity is described in more detail in Appendix A2.

The signal that remains after removing cardiac activity shows both positive and negative deflections. To obtain a continuous positive signal of general muscle activity, the root-mean-square (RMS) is calculated [51], which is depicted in the third graph of Figure 3.14. The RMS of an EMG signal is a measure of its intensity or power and is calculated by taking the square root of the mean of the squares of the individual sample value:

$$RMS = \sqrt{\frac{1}{n}\sum_{n} x_n^2} \quad (3.2)$$

where x_n represent the values of the EMG signal with n the number of samples used per step. The

RMS was calculated for every sample in the dataset with a moving window of 0.1 seconds.

The initial cardiac artifact removal did not completely eliminate all cardiac artifacts, so a second cardiac artifact removal process was implemented. This method also used template subtraction, but with two templates. For each artifact, the first template was created by taking the average of 21 artifacts: the current artifact, 10 artifacts before it and 10 artifacts after it. The correlation of each artifact with the first template was calculated, and only artifacts with a correlation higher than 0.5 were used to construct the second template. The offset and amplitude of the second template were then adjusted using linear regression. The adjusted template was subtracted from the RMS signal, and the RMS signal was resampled from 500 Hz to 50 Hz.

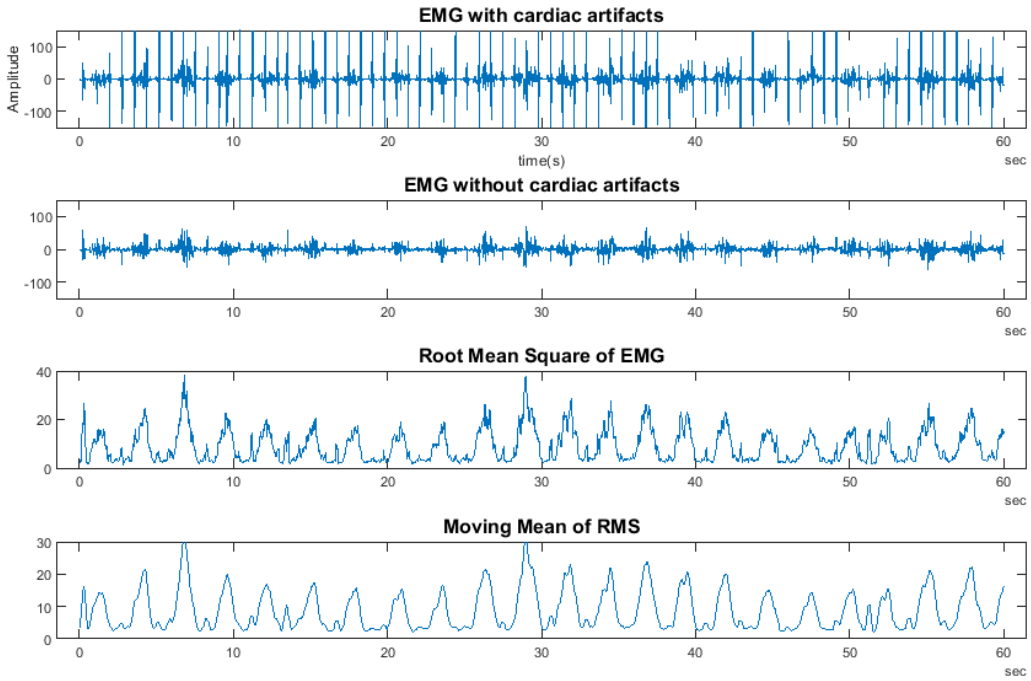


Figure 3.13: Removal of cardiac activity from the EMG signal

3.3.2 Pre-processing part 2

In order to optimize the interpretation of EMG, it is necessary to remove both cardiac interference and motion artifacts from the signal. These motion artifacts can be difficult to predict and vary significantly in shape and magnitude, making them challenging to remove [51, 53]. To ensure that only reliable features are obtained, several steps are taken, which can be seen in Figure 3.14.

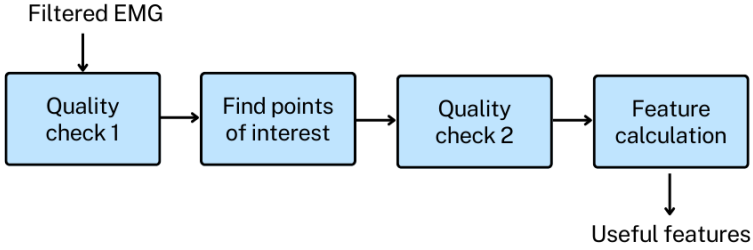


Figure 3.14: Flowchart of EMG preprocessing part 2

Quality check 1
To remove data with high levels of missing values or motion artifacts, the following steps were taken:

- For each minute of data, it was determined whether more than half of the data points were missing. If this was the case, the minute was discarded.
- A check was performed to identify any minutes with abnormal high amplitudes, defined as differences between the highest and lowest point above 300. If such a minute was found, it was discarded.
- A check was performed to identify minutes with more than 10% missing values or more than 10 consecutive missing values. If such a minute was found, it was discarded.

The last criterion relates to the interpolation of the data, which occurs at a later stage. Interpolation becomes unreliable when there are too many missing values. By discarding data with high levels of missing values or motion artifacts, the goal is to ensure that the resulting features are based on reliable data.

Points of interest
Following the initial quality check, the points of interest are identified and can be seen in Figure 3.15. For each breath, the maximal muscle activity, as well as the start and end of the breath, are determined. The start or end of a breath is defined as the midpoint between the end of the muscle activity of the previous breath and the start of the muscle activity of the current breath.

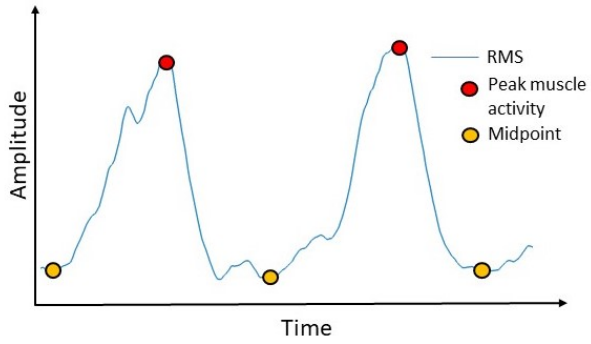


Figure 3.15: Points of interest in EMG signal where the red circles indicate the breathing peaks and the yellow ones indicate the midpoints between breaths.

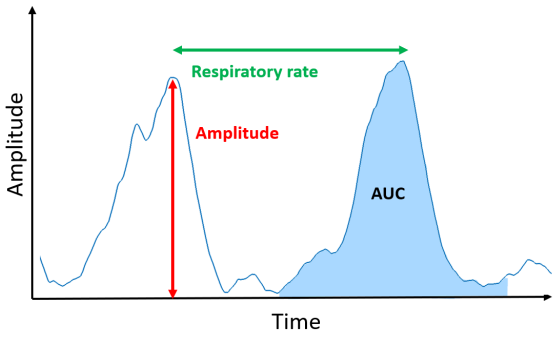


Figure 3.16: Features of the EMG signal.

Second quality check
The second quality check determines whether at least half of the waves in the minute are useful breaths. Breaths with an amplitude less than 0.5 times the median amplitude or above 1.5 times the median amplitude for that minute are excluded. In addition, breaths with start or end points that are located at the same point as the maximum muscle activity are also considered not useful, as the points had not been accurately determined. From the remaining breaths, the prominence of the peaks is calculated by measuring how much the peaks stand out due to its intrinsic height and its location relative to other peaks [54]. By dividing this prominence by the noise in the minute, a signal-to-noise ratio is obtained. Noise is defined as the average of the differences between consecutive samples. Minutes with a signal-to-noise ratio lower than 0.60 are discarded. The distance between breaths is also calculated. If the variation in the distance between the peaks is too large, these minutes are also excluded.

Feature calculation
Four features were derived from the EMG signals: respiratory rate, amplitude, area under the curve per breath, and breathing intensity, which can be seen in Figure 3.16.

Respiratory rate
The peak-to-peak distance of the EMG waveforms was used to calculate the respiratory rate, with adjustments made to account for any incorrectly identified peaks. Sometimes, the breathing peaks were widely spaced or very close together, resulting in unrealistic respiratory rates. For this reason, respiratory frequencies below 8 or above 40 were removed.

Amplitude and Area under the Curve
To accurately compare the amplitudes and areas under the curve between patients, the RMS signal was first normalized. This is done because the amplitudes of the signals can vary greatly between patients due to factors such as differences in the amount of tissue between the electrodes and the diaphragm or variations in the placement of the electrodes. Normalization allows for more meaningful comparison of the signals and was performed in the following way:

$$x_{normalized} = \frac{x_n - x_{min}}{x_{max} - x_{min}} \tag{3.3}$$

where x_{min} is the average of the lowest 25% in the dataset and x_{max} is the average of the highest 25% in the dataset. To minimize the influence of any potential artifacts in the data, it was decided not to use the absolute highest and lowest values as x_{min} and x_{max}. The amplitude was subsequently determined by finding the highest absolute value of the EMG waveform per breath. The area under the curve is calculated by taking the integral for each breath from beginning to end of breath.

Breathing Intensity
To accurately evaluate the effort the patient expends to breathe, the breathing intensity is calculated by multiplying the area under the curve with the respiratory rate. It was not reliable to calculate the total area under the curve for a given time interval, as this would also include the area of any motion artifacts present. To avoid this issue, the areas under the curve of the reliable respiratory peaks are averaged and multiplied by the average respiratory rate for that time period. This method provides a more accurate representation of the patient's breathing intensity.

3.4 Trend analysis

3.4.1 Inclusion criteria

In some cases, the data quality was poor, making it difficult to determine the features or the data was only collected for a short period of time. For the PPG data, a patient was included in the trend analysis if the measurement contained at least 4 hours of data and if 20% of the available waves were classified as good waves. For the EMG analysis, a patient was included if there was at least 4 hours of data available and if the data consisted of a minimum of 60% good waves. The lower threshold for PPG data was chosen, because there are more waves per minute from which features can be extracted, requiring less data for a reliable average. These cut-off values were determined subjectively, based on the belief that reliable trend analysis was still possible with these criteria.

Already during initial analysis, it occurred that the sample size of the EMG group was relatively small. As a result, a secondary analysis was conducted with relaxed criteria, using only basic trend features, such as the maximum, minimum, difference, median, standard deviation and coefficient of variation. The inclusion criteria were set to require at least 100 good minutes of data over a minimum of two hours.

3.4.2 Median feature time-series

Trends across time were then determined for all identified features on both the PPG and EMG data. Prior to this analysis, the median of each feature was calculated for every minute over a 10-minute interval in order to eliminate remaining outliers and provide a reliable measure of the data over time. To calculate the median for this 10-minute interval, a minimum of 80 data points (approximately 10% of the entire window) of the relevant feature was required. If there were insufficient data points, the median was deemed unreliable. The transformation from the raw feature to a median smoothed feature is depicted in the top two figures of Figure 3.18. Following this, several trend features were identified, as displayed in Figure 3.17.

3.4.3 Trend features

Based on the calculated median over time from the previous step, basic trend features were calculated. These features included the minimum, maximum, delta, median, standard deviation and coefficient of variation. The minimum and maximum were determined as the average of the lowest 2% and the highest 2% of the data, respectively, with the delta representing the difference between these values. Subsequently, more complex trend features were calculated, including the number of trend changes, the ratio of time spent trending upwards to trending downwards, the longest trend, the steepest trend and the overall trend. The following paragraphs will provide further explanation of these more complex trend features.

Number trend changes

To quantify the level of instability, the number of trend changes can be used as a feature. This can be calculated by counting the number of local extreme values, as each extreme value signifies a change in the trend of the vital sign (e.g. there is an increasing trend before a local maximum and a decreasing trend after it). To determine the number of trend changes, all local maxima and minima must be identified. However, for this to be feasible, the median of the feature must be a continuous time-series. To achieve this, the feature time-series is interpolated if it not already continuous. If it contains large gaps of missing data, interpolation may not be reliable. In these cases, the data is checked for gaps larger than one hour. If such gaps are present, the two pieces of data are joined together if the following criteria are met: the data before and after the gap must be at least 60 minutes in length and contain a maximum of 50% missing values. If these criteria are not satisfied, the largest portion of available data (either before or after the gap) was used for further analysis. Any remaining missing values are interpolated if the following conditions were met: the pieces of data before and after the gap must be at least 30 minutes in length and contain a maximum of 50% missing values. After this, the continuous signal is smoothed using a moving mean filter with a window size of 5 samples, and local minima and maxima are determined. An example of these minima and maxima can be seen in the third plot of Figure 3.18. The sum of the minima and maxima was then divided by the number of minutes over which it was calculated, allowing for proper comparison between patients.

Ratio feature up and down

By using previously determined minima and maxima, it is possible to calculate the time period that the feature was increasing or decreasing during the complete measurement. The ratio of these time periods can then be calculated. In addition, it is also useful to assess the magnitude of upward and downward fluctuations in the features. As such, the ratio of the feature's amplitude in increasing and decreasing states is also determined.

Longest slope

The slope and duration of the longest trend in the feature are then calculated. It is possible that the process of joining pieces of data may have artificially created the longest trend. These pieces of data were considered as separate in this evaluation.

Steepest slope
To identify the steepest slope, the first derivative of the feature was taken. The largest derivative was then identified and verified to have a positive slope for a period of at least six minutes (three minutes before and three minutes after). This was done to ensure that a brief, rapid increase in the feature was not included as the steepest slope. It was also taken into account that the steepest slope may have been artificially created by joining pieces of data and corrective measures were implemented.

Slope overall measurement
The last feature is determined by calculating the slope across the entire measurement using linear regression to fit a first-order polynomial to the feature's time-series, without joining segments or performing any interpolation. This can be observed in the fourth graph of Figure 3.18.

Trend features
Minimum and maximum
Delta (maximum- minimum)
Median
Standard Deviation
Coefficient of Variation
Number of trend changes
Ratio time up/down
Ratio amount up/down
Longest trend - slope - duration
Slope steepest trend
Slope entire measurement

Figure 3.17: Trend Features

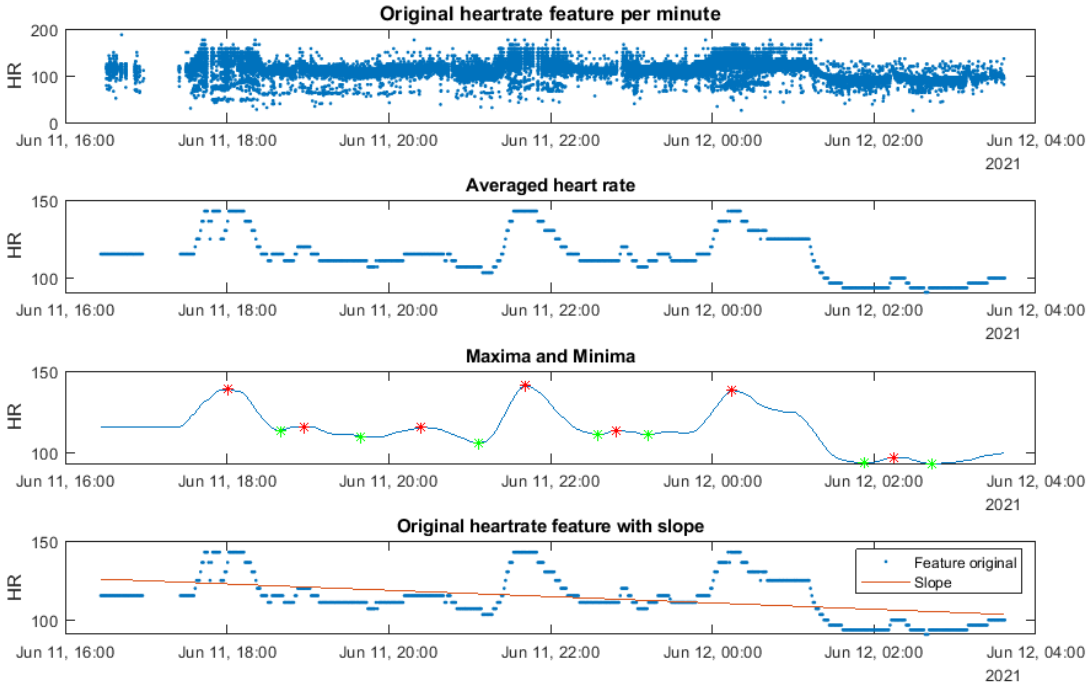


Figure 3.18: From raw feature to trend features.

3.5 Statistical analysis

In order to investigate the differences in vital and demographic characteristics between the deteriorated and non-deteriorated groups, statistical analysis was conducted using Stata (StataCorp, 2017). The categorical variables were analyzed using the Pearson's Chi-square test, while the continuous variables were analyzed using the Mann-Whitney U test. In order to determine whether there were significant differences in the trend features of the EMG and PPG data between the two groups, a Mann-Whitney U test was also performed. This non-parametric test was chosen due to the non-normal distribution and lack of pairing in the data. A p-value of 0.05 or less was initially considered as significant. However, given the large number of features in the PPG data, the multiple comparisons problem was taken into consideration, and the p-value was corrected using the False Discovery Rate (FDR) method as described in [55].

As a result of normalizing the PPG signals, the amplitude is changed and thus it becomes unfeasible to compare amplitude related features between groups. As a result, the systolic and diastolic peak amplitude features (minimum, maximum, mean, median) were excluded from analysis.

3.6 Machine-learning Models

Given the limited size of the EMG data, the focus was on developing predictive models using PPG data and vital signs. Various models, including Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Decision Tree, and Random Forest, were applied to the data using different combinations. This included vital signs only, a subset of PPG features (including all initially significant and the 15 most effective ones), and a combination of both. Using the 15 most effective features can reduce overfitting, improve model efficiency and generalization performance, and eliminate the impact of irrelevant features. The performance of these models was also compared against the qSOFA scoring system, which is a bedside tool to predict patients at risk for poor outcomes outside of the ICU based on their blood pressure, respiration rate, and Glasgow Coma Scale. Although qSOFA was not explicitly designed for predicting sepsis, it is of interest to examine the impact of adding vital parameters on prediction performance.

Imputation

In the vital parameter data, respiration rate, temperature and Glasgow Coma Scale (GCS) were frequently missing. These variables are usually not measured unless there is a specific need, so it was assumed that their values were normal and they were imputed as 14, 37.45, and 15, respectively. The values for respiration rate and GCS are based on what is considered normal for a resting individual [56, 57]. The specific value for temperature was selected based on the median of the available data. For the patients where blood pressure values were not recorded, the systolic blood pressure and diastolic blood pressure were imputed using the KNN imputation method with 5 neighbors. The mean arterial pressure (MAP) was subsequently calculated from these values [58].

Data augmentation

In the dataset, the classes were not equally represented, which was already expected. To address this imbalance, Synthetic Minority Oversampling Technique (SMOTE) was applied to the training set [59]. SMOTE was designed to oversample the minority class in imbalanced data sets, allowing for training on equally represented data.

Feature selection

The selection of features for the models was carried out using the filter method. The procedure began with the use of a statistical test known as the Mann-Whitney U test to determine which features showed a significant difference between the two groups. The final set of features, consisting of the 15 most effective ones, was then created by merging the features that passed the Mann-Whitney U test, removing those with high correlation, and selecting the features considered important based on the results of logistic regression and random forest models.

Validation

To assess the generalization abilities of the classifiers, stratified k-fold cross validation was used [60]. Stratified k-fold cross validation is a method of evaluating the performance of a model by dividing the data into k folds, training the model on k-1 folds, and using the remaining fold for internal validation. In this study, k was set to 5 folds to avoid having a validation set that is too small. This technique is particularly useful for imbalanced data sets as it ensures that each fold has a representative proportion of different classes. This approach allows for a more accurate estimate of model performance, as it accounts for the variability that may arise due to the specific samples selected for training and evaluation. To measure the performance of the classifiers, accuracy, precision, recall, AUC and the F1 score are calculated.

4 RESULTS

In this chapter the outcomes of the study are presented. The first step is to describe the study population, followed by a presentation of the vital and demographic parameters. Then, the results of the statistical analysis will be presented. The predictive models are then presented, starting with a model that only uses vital signs and then evaluating the addition of PPG features.

4.1 Population

In this study, patients who visited the ED at the UMCG between 9am and 11pm from March 2021 to October 2022 were included. The methodology for patient inclusion and grouping is presented in two flow diagrams, with the flowchart for the PPG data shown in Figure 4.1 and the flowchart for EMG data in Figure 4.2. Out of the 235 patients with PPG data, 189 patients with confirmed infections were used for trend analysis and 24 experienced deterioration. A diaphragmatic EMG measurement was carried out in 61 patients, with 25 eligible for trend analysis and 1 showing deterioration. As a result, a basic trend analysis with other inclusion criteria was performed on 38 patients, with 6 experiencing deterioration.

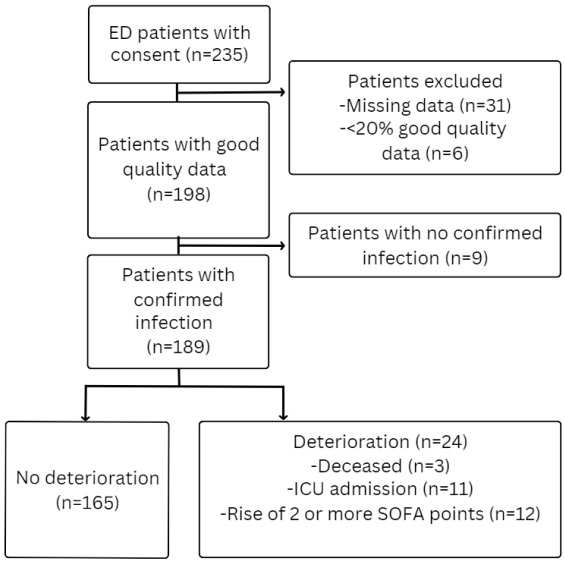


Figure 4.1: PPG patient selection flow.

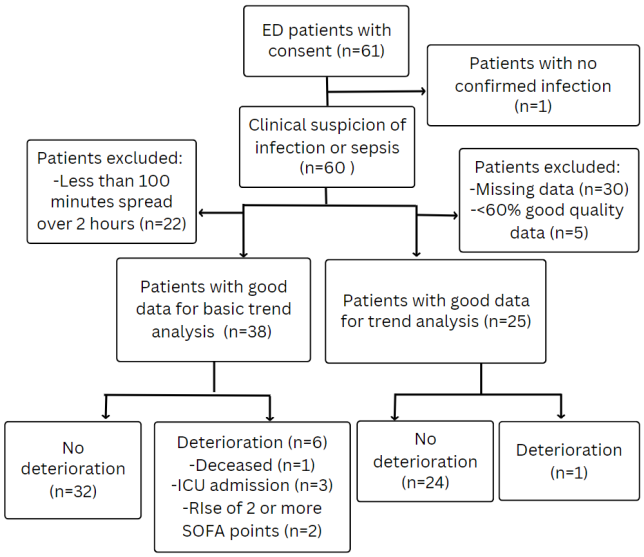


Figure 4.2: EMG patient selection flow.

4.2 Vital signs and demographics

The vital signs of the included subjects collected at triage are described in Table 4.1. The study sample consists of 120 male and 69 female patients with a median age of 66 [59-72]. Statistical tests were conducted to compare the vital signs between the non-deteriorated and deteriorated group. Results showed that the deteriorated group had a lower median diastolic blood pressure (66 [61-77] versus 76 [67-86]), a higher median respiration rate (25 [16-27] versus 20 [14-24]) and a lower Glasgow Coma Score at triage. Figure 4.3 shows the correlations between the variables and shows strong positive correlations between blood pressure readings and a negative correlation between respiration and oxygen saturation.

The SOFA scores were calculated during the first 24 hours after admission (day 0), after 24 hours (day 1), and after 48 hours (day 2). The median day 0 SOFA scores for all patients were 2.0 [0.0-3.0], 1.0 [0.0-2.0] for day 1, and 0.0 [0.0-1.0] for day 2. The day 0, 1, and 2 SOFA scores for the deteriorated group were 3.0 [0.75-5.0], 2.5 [1.75-3.25], and 3.0 [1.0-4.0], respectively, while the corresponding scores for the non-deteriorated group were 2.0 [0.0-3.0], 0.0 [0.0-1.0], and 0.0 [0.0-1.0].

Table 4.1: Vital and demographic parameters measured at triage for the total study population and per group.

	All (n=189)	**Non-deteriorated (n=24)**	**Deteriorated (n=165)**	**p-value**
Variable				
Age [years]	66 [59 - 73]	66 [59 - 72]	67 [62 - 75]	0.300
Male gender [%]	120 (63.5)	105 (63.6)	15 (62.5)	0.910
Heart rate [bpm]	100 [90 - 115]	100 [90 - 112]	100 [86 - 128]	0.600
SBP [mmHg]	125 [114 - 145]	127 [115 - 146]	119 [107- 134]	0.120
DBP [mmHg]	75 [65 - 86]	76 [67 - 86]	66 [61 - 77]	0.028*
MAP [mmHg]	93.0 [83.7 - 103]	93.3 [84.7 - 104]	89.3 [75.7 - 96.7]	0.050
SpO2 [%]	96 [94 - 98]	96 [94 - 98]	96 [96 - 98]	0.390
Respiration rate [bpm]	20 [14 - 25]	20 [14 - 24]	25 [16 - 27]	0.044*
Temperature [degrees Celcius]	37.4 [36.7 - 38.2]	37.4 [36.7 - 38.2]	37.4 [36.6 - 38.1]	0.750
GCS - 12 - 13 - 14 - 15	 1 (0.5%) 4 (2.1%) 8 (4.2%) 176 (93.1%)	 0 (0.0%) 3 (1.8%) 6 (3.6%) 156 (94.5%)	 1 (4.2%) 1 (4.2%) 2 (8.3%) 20 (83.3%)	0.042*

Continuous and ordinal variables are presented as median and interquartile ranges [IQR], and are compared using the Mann-Whitney U test. Categorical variables are expressed as proportions and compared using Pearson's Chi-square test. SBP= systolic blood pressure, DBP= diastolic blood pressure, MAP= mean arterial pressure, SpO2= oxygen saturation, GCS= glasgow coma scale. * indicates a significant p-value.

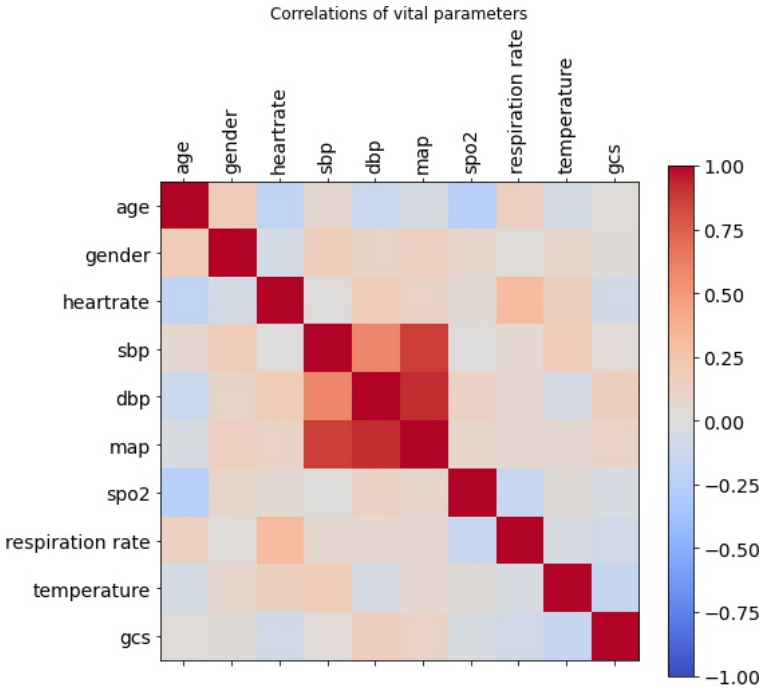


Figure 4.3: The correlation plot of the vital parameters shows the strength and direction of the relationships between the variables. The scale ranges from total positive correlation (+1) to no correlation (0) to total negative correlation (-1), with red indicating positive correlation and blue indicating negative correlation.

4.3 PPG features

In this study, the Mann-Whitney U test was conducted to determine significant differences between patients with deterioration and those without. Out of 210 PPG features, 55 were initially found to be statistically significant at a p-value of 0.05, without accounting for the multiple comparisons problem. After applying the False Discovery Rate correction, only one feature, the crest time- ratio amount, was deemed significant at an adjusted p-value of 1.69e-04.

Table 4.2 shows the median and interquartile ranges of the top 15 PPG features selected after removing highly correlated features such as pulse interval and heart rate. The results show that among patients who deteriorated, there was a decrease in the crest time- ratio amount, median stroke volume, and an increase in the variability of several PPG features. This increase in variability was reflected by a higher number of trend changes in diastolic peak amplitude, reflection index, and infliction point area, as well as an increase in the coefficient of variation for the diastolic peak amplitude.

Furthermore, several heart rate related features (6 out of 15) were found to be related to deterioration. Patients who deteriorated had a positive longest slope in heart rate compared to a negative slope in those who did not deteriorate, a lower median pulse interval, a higher heart rate- ratio amount, a lower minimum in the Peak to Peak interval, and an increased delta in heart rate.

Finally, Table 4.2 shows that the majority of the features (9 out of 15) are more complex trend features, such as ratio amount, ratio time, number of trend changes, and longest slope. The tables of all PPG features, including the values for both groups and their p-values, can be found in Appendix A.3. A comprehensive summary of trend features and their associated p-values can be found in Appendix A.4.

4.4 EMG features

A Mann-Whitney U test was performed to evaluate the statistical significance in EMG features between the deteriorating and non-deteriorating group. The results, as illustrated in Table 4.3, depict the median and interquartile ranges of all EMG features, along with their corresponding p-values. The results indicate that four features are statistically significant, with three of these features being related to respiratory rate. The group experiencing deterioration had a lower delta in respiratory rate, indicating a smaller difference between the minimum and maximum respiratory rate. Additionally, the group experiencing deterioration had a lower standard deviation and coefficient of variation within the respiratory rate, indicating a reduced degree of variability in respiratory rate. Lastly, the deteriorated group had a higher minimum of the area under the curve. No statistically significant differences were identified in the amplitude and breathing intensity features.

Table 4.2: Top 15 PPG trend features for deteriorated and non-deteriorated group.

	Deteriorated (n=24)	**Non-deteriorated (n=165)**	**p-value**
Trend feature			
Crest Time - Ratio Amount (-)	0.87 [0.61 - 1.00]	1.05 [0.83 - 1.28]	<0.001*
Stroke Volume - Median (a.u)	9.81 [9.00 - 12.88]	12.81 [11.00 - 14.46]	<0.01*
Diastolic peak amplitude - Number of trend changes (-)	0.06 [0.05 - 0.07]	0.05 [0.04 - 0.05]	<0.01*
Reflection Index - number of trend changes (-)	0.06 [0.05 - 0.06]	0.05 [0.04 - 0.06]	<0.01*
Heart rate - Longest slope (-)	0.10 [-0.06 - 0.21]	-0.06 [-0.10 - 0.05]	<0.01*
Delta Time - Median (s)	0.22 [0.22 - 0.26]	0.26 [0.24 - 0.26]	<0.01*
Pulse Interval - Median (s)	0.62 [0.55 - 0.72]	0.70 [0.62 - 0.82]	0.01*
Heart rate - Ratio amount (-)	1.14 [0.89 - 1.76]	0.90 [0.73 - 1.08]	0.01*
Crest Time - Ratio time (-)	0.79 [0.43 - 1.17]	1.14 [0.69 - 1.76]	0.01*
Pulse Interval - Longest duration (min)	89 [55 - 105]	110.5 [81.5 - 160.5]	0.01*
Peak to Peak interval - Minimum (s)	0.54 [0.47 - 0.58]	0.60 [0.53 - 0.66]	0.01*
Diastolic peak amplitude - Coefficient of Variation (-)	0.95 [0.87 - 1.05]	0.88 [0.82 - 0.97]	0.01*
Infliction Point Area - Number of trend changes (-)	0.06 [0.04 - 0.06]	0.05 [0.04 - 0.05]	0.01*
Acceleration Plethysmography - Number of trend changes (-)	0.05 [0.04 - 0.06]	0.05 [0.04 - 0.05]	0.01*
Heart rate - Delta (bpm)	29.78 [26.15 - 41.67]	25.81 [18.74 - 34.35]	0.01*

The features are expressed as median and the interquartile ranges [IQR], and are compared using the Mann-Whitney U test. * indicates a significant p-value.

Table 4.3: EMG basic trend features for deteriorated and non-deteriorated group.

	Deteriorated (n=6)	**Non-deteriorated (n=33)**	**p-value**
Trend feature			
Respiration rate	.....	.....	.....
- Minimum	17.75 [15.50- 19.52]	17.91 [15.14 - 19.49]	1
- Maximum	23.98 [20.89 - 32.68]	28.47 [23.83 - 31.48]	0.448
- Delta	5.78 [2.92 - 6.30]	9.57 [7.10 - 11.76]	0.021*
- Mean	20.08 [17.88 - 26.73]	22.17 [20.15 - 25.51]	0.599
- Median	19.81 [17.54 - 26.79]	22.39 [19.94 - 25.99]	0.599
- Std	1.37 [0.71 - 1.59]	2.02 [1.61 - 2.02]	0.019*
- CV	0.07 [0.05 - 0.08]	0.10 [0.08 - 0.12]	0.028*
Area under the Curve	.....	.....	.....
- Minimum	20.61 [17.82 - 35.90]	15.86 [10.44 - 19.90]	0.031*
- Maximum	73.96 [30.66 - 105-.18]	88.66 [59.49 - 142.36]	0.403
- Delta	44.06 [11.69 - 69.28]	59.75 [39.60 - 124.95]	0.251
- Mean	32.69 [25.14 - 61.28]	37.00 [26.64 - 44.96]	0.800
- Median	27.83 [23.84 - 55.39]	30.15 [25.41 - 40.40]	0.953
- Std	9.73 [2.58 - 17.73]	14.31 [10.08 - 24.16]	0.267
- CV	0.24 [0.11 - 0.30]	0.48 [0.27 - 0.61]	0.070
Amplitude	.....	.....	.....
- Minimum	0.37 [0.34 - 0.72]	0.31 [0.22 - 0.43]	0.155
- Maximum	1.34 [1.05 1.83]	1.34 [1.04 - 1.79]	0.922
- Delta	0.81 [0.22 - 1.46]	0.85 [0.63 - 1.53]	0.496
- Mean	0.86 [0.57 - 1.02]	0.73 [0.52 - 0.88]	0.425
- Median	0.82 [0.54 - 0.98]	0.66 [0.47 - 0.84]	0.360
- Std	0.17 [0.05 - 0.31]	0.19 [0.14 - 0.34]	0.381
- CV	0.22 [0.09 - 0.28]	0.36 [0.21 - 0.54]	0.106
Breathing intensity	.....	.....	.....
- Minimum	527 [365- 753]	323 [251 - 451]	0.090
- Maximum	1718 [1215 - 2136]	1821 [1409 - 2864]	0.496
- Delta	1157 []343 - 1450]	1431 [822 - 2535]	0.321
- Mean	917 [681 - 1140]	785 [620 - 1039]	0.599
- Median	807 [631 - 1056]	720 [513 - 914]	0.572
- Std	228 [80-355]	306 [192 - 547]	0.220
- CV	0.26 [0.10 - 0.30]	0.45 [0.28 - 0.59]	0.059

The features are expressed as median and the interquartile ranges [IQR], and are compared using the Mann-Whitney U test. * indicates a significant p-value.

4.5 Predictive models

In order to explore the potential for developing a predictive model of deterioration, first a model was constructed using only vital signs and demographics. Next, models were developed using PPG features and the combination of both.

4.5.1 Model vital signs and demographics

The validation results of the different classifiers based on vital parameters and demographics are presented in Figure 4.4. The results indicate that the classifiers show a minimal improvement over chance and perform slightly better than the qSOFA metric. However, the magnitude of this improvement is small. A summary of the performance measures for all models is presented in Table 4.4, which shows the low training, validation and the other performance scores. Despite the overall poor performance of the classifiers, the logistic regression model showed the best results with an AUROC score of 0.63 and a F1 score of 0.447.

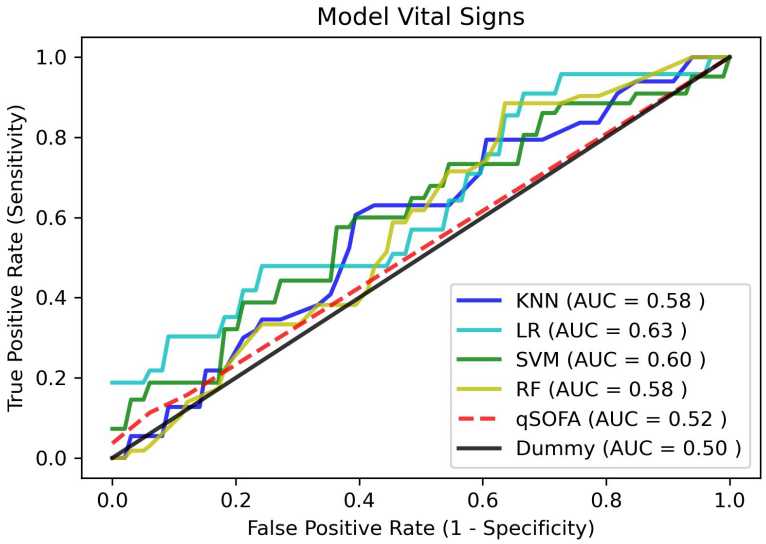


Figure 4.4: Predicting deterioration based on vital signs and demographics.

Table 4.4: Results of the validation experiments with the vital parameters, presented as the mean and standard deviation of the 5-fold cross-validation.

Classifier	Training	Validation	F1 Score	ROC-AUC	Sensitivity	Specificity
KNN	0.68 (0.03)	0.57 (0.10)	0.55 (0.16)	0.58 (0.15)	0.56 (0.20)	0.58 (0.05)
LR	0.71 (0.03)	0.53 (0.16)	0.45 (0.25)	0.63 (0.14)	0.42 (0.25)	0.64 (0.11)
SVM	0.46 (0.18)	0.54 (0.12)	0.46 (0.18)	0.60 (0.13)	0.41 (0.20)	0.67 (0.08)
RF	1.0 (0.0)	0.45 (0.02)	0.06 (0.07)	0.58 (0.134)	0.04 (0.05)	0.88 (0.05)
qSOFA	-	-	0.16 (0.07)	0.52 (0.08)	0.10 (0.14)	0.94 (0.04)

Training and validation scores are measured by accuracy. KNN= K-Nearest Neighbors, LR= Logistic Regression, SVM= Support Vector Machine, RF= Random Forest. qSOFA= Quick Sequential Organ Failure Assessment. ROC-AUC= Receiver Operating Characteristic - Area Under the Curve.

4.5.2 Model PPG features

As previously mentioned, 55 features were initially found to be statistically significant at a p-value of 0.05, and the first PPG model was developed incorporating these features. The results, as shown in Figure 4.5, demonstrate the predictive capacity of the PPG features, with all classifiers performing superior to chance. Table 4.5 reveals that the KNN model performs best, achieving an AUROC score of 0.77 and a F1 score of 0.66. The LR, SVM and RF models show high training scores and lower validation scores, indicating overfitting. The high degree of correlation among many of the model's features can be observed in Figure A.3 in Appendix A.5.

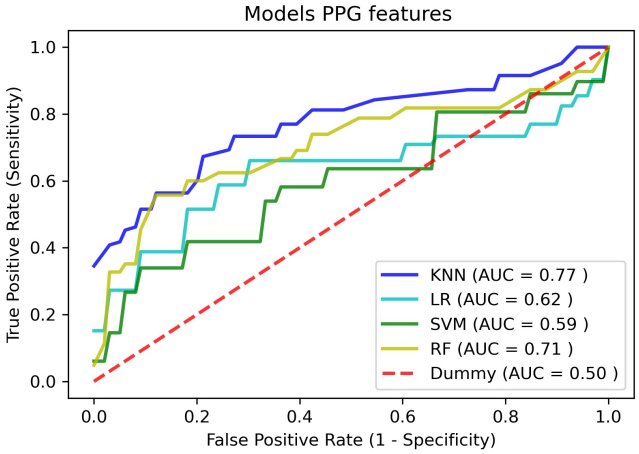


Figure 4.5: Predicting deterioration using all significant features.

Table 4.5: Results of the experiments with all significant PPG features, presented as the mean and standard deviation of the 5-fold cross-validation.

Classifier	Training	Validation	F1 Score	ROC-AUC	Sensitivity	Specificity
KNN	**0.80 (0.02)**	**0.70 (0.12)**	**0.66 (0.23)**	**0.77 (0.20)**	**0.72 (0.32)**	**0.69 (0.14)**
LR	0.95 (0.03)	0.68 (0.04)	0.61 (0.08)	0.63 (0.11)	0.52 (0.12)	0.84 (0.04)
SVM	0.97 (0.01)	0.60 (0.07)	0.48 (0.11)	0.53 (0.08)	0.38 (0.11)	0.83 (0.07)
RF	1.0 (0.0)	0.63 (0.14)	0.41 (0.29)	0.71 (0.22)	0.32 (0.26)	0.95 (0.04)

Training and validation scores are measured by accuracy. Best results are printed in bold. KNN= K-Nearest Neighbors, LR= Logistic Regression, SVM= Support Vector Machine, RF= Random Forest. ROC-AUC= Receiver Operating Characteristic - Area Under the Curve.

4.5.3 Model PPG subset features

A subsequent model was developed using the 15 best features found through statistical testing, correlation plot analysis, and feature importance assessment using LR and RF analysis. The correlation plot of the selected features can be seen in Figure A.4 in Appendix A.5. The results of the classifiers using these 15 features are shown in Figure 4.6 and Table 4.6. The KNN model had the best performance with the highest F1 score (0.73), AUROC score (0.79), and sensitivity (0.73). All models showed improvement, especially LR, SVM, and RF, with reduced overfitting compared to the previous model.

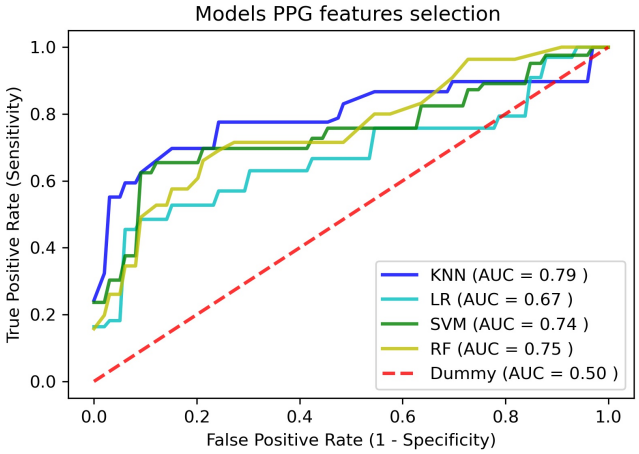


Figure 4.6: Predicting deterioration based on best fifteen features.

Table 4.6: Results of the experiments with the selection of 15 PPG features, presented as the mean and standard deviation of the 5-fold cross-validation.

Classifier	Training	Validation	F1 score	ROC-AUC	Sensitivity	Specificity
KNN	**0.77 (0.04)**	**0.76 (0.10)**	**0.73 (0.14)**	**0.79 (0.12)**	**0.73 (0.22)**	**0.78 (0.08)**
LR	0.82 (0.02)	0.66 (0.12)	0.61 (0.18)	0.67 (0.17)	0.57 (0.21)	0.75 (0.12)
SVM	0.84 (0.03)	0.76 (0.07)	0.74 (0.07)	0.74 (0.10)	0.70 (0.10)	0.81 (0.11)
RF	1.0 (0.0)	0.69 (0.15)	0.54 (0.32)	0.75 (0.14)	0.46 (0.29)	0.93 (0.05)

Training and validation scores are measured by accuracy. Best results are printed in bold. KNN= K-Nearest Neighbors, LR= Logistic Regression, SVM= Support Vector Machine, RF= Random Forest. ROC-AUC= Receiver Operating Characteristic - Area Under the Curve.

4.5.4 Model PPG features with vitals

The objective of of this last analysis is to evaluate the effect of adding the vital signs to the previous PPG model. To accomplish this, the performance of the model using both the subset of PPG features and vital parameters is compared against that of the previous model. As demonstrated in Figure 4.7 and Table 4.7, the incorporation of vital signs and demographics resulted in an improvement of the performance of all models. KNN has the highest F1, AUROC score and sensitivity among the four classifiers.

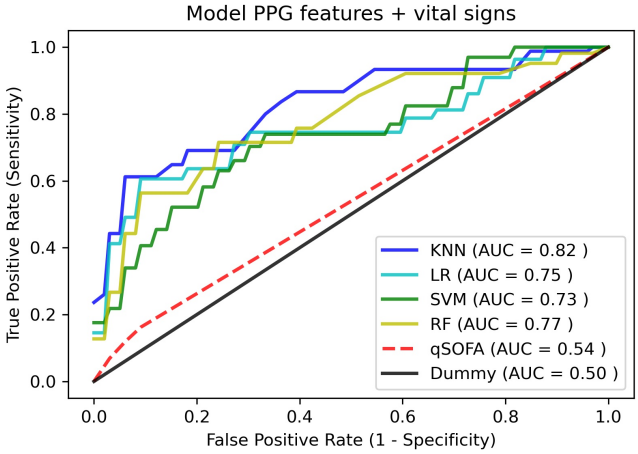


Figure 4.7: Predicting deterioration based selection PPG features and vital signs

Table 4.7: Results of the experiments with the selection of 15 PPG features and the vital signs, presented as the mean and standard deviation of the 5-fold cross-validation.

Classifier	Training	Validation	F1 score	ROC-AUC	Sensitivity	Specificity
KNN	**0.81 (0.01)**	**0.74 (0.04)**	**0.73 (0.07)**	**0.82 (0.07)**	**0.71 (0.15)**	**0.76 (0.09)**
LR	0.92 (0.02)	0.73 (0.16)	0.68 (0.22)	0.75 (0.11)	0.64 (0.25)	0.82 (0.10)
SVM	0.93 (0.02)	0.65 (0.15)	0.51 (0.25)	0.73 (0.13)	0.43 (0.27)	0.86 (0.05)
RF	1.0 (0.0)	0.63 (0.07)	0.44 (0.15)	0.77 (0.06)	0.31 (0.12)	0.95 (0.02)

Training and validation scores are measured by accuracy. Best results are printed in bold. KNN= K-Nearest Neighbors, LR= Logistic Regression, SVM= Support Vector Machine, RF= Random Forest. ROC-AUC= Receiver Operating Characteristic - Area Under the Curve.

5 DISCUSSION

This study demonstrates that combining trends in PPG waveforms with vital signs can effectively predict patient deterioration. The K-nearest Neighbor (KNN) model showed the best performance with an AUROC score of 0.82 and an F1-score of 0.73. These findings suggest the potential utility of continuous PPG waveforms at the ED and nursing wards and warrant further investigation towards the development of a clinical tool. Unfortunately, no trend analysis could be performed on the EMG data due to limitation in signal quality and the number of patients.

5.1 Interpretation results

5.1.1 Vital signs and demographics

The correlation analysis of vital signs and demographics revealed strong positive associations between blood pressure readings and a negative association between respiration and oxygen saturation, which is in line with expectations. The observation of reduced diastolic blood pressure, increased respiration rate, and a lower GCS score among patients experiencing deterioration can be attributed to their presenting in a compromised state due to sepsis symptoms. This would have resulted in decreased cardiac output and perfusion, contributing to the decline in diastolic blood pressure. In a compensatory effort, the body increases the respiration rate to improve oxygen delivery.

Additionally, it was observed that the SOFA scores of patients who did not deteriorate decreased over a 48-hour period, which is a logical outcome, as hospitalization improved their health status. The deteriorated group did not show changes in their SOFA scores. This could be attributed to the fact that this group also includes patients who passed away or were admitted to the Intensive Care Unit (ICU). In these circumstances, it is possible that further organ failure did not occur.

5.1.2 PPG trend features

After correction for False Discovery Rate, only the Crest time - ratio amount feature was found to be significant. Crest time, which refers to the duration of the systolic phase, is defined as the interval between the initiation of the systolic peak and the peak itself. It reflects the duration required for blood to be expelled from the heart and reach the peripheral artery. The observation of a lower ratio amount in the deteriorated group suggests that the bigger part of the measurement the crest time was decreasing over time, indicating that the cardiac ejection is occurring at a faster rate. This could be a compensatory mechanism in response to the decreased cardiac output or decrease in perfusion, as supported by the lower median cardiac output (Table A.15) and stroke volume (Table A.14) in the deteriorated group, as presented in Appendix A.3.

The second most effective feature was the median stroke volume, which was lower in the deteriorated group. This aligns with literature on early sepsis, as a low stroke volume is reported at the initial phase of sepsis due to decreased cardiac preload resulting from high vascular permeability and vasodilation [61]. The compensatory tachycardia is often insufficient to maintain adequate cardiac output during the early phase of sepsis. This is supported by the higher mean heart rates (Table A.11) and reduced cardiac output (Table A.15) in the deteriorated group, which can be seen in Appendix A.3.

Additionally, the deteriorated group demonstrated a higher number of trend changes in various PPG features, including diastolic peak amplitude, reflection index, inflection point area, and acceleration plethysmography. This higher degree of variability among these features is in line with the findings of Bloch et al. [37], who demonstrated the capability of accurately predicting the onset of sepsis in the ICU based on variations in physiological parameters. However, it is important to keep in mind that the variability may not only be a result of underlying pathophysiology, but also due to interventions given to patients deteriorating, such as medication, which can alter the PPG signal and increase variability in PPG features.

Furthermore, several heart rate features were found to be related to deterioration. Analysis revealed that the deteriorated group had higher mean, median, delta and maximum heart rate (Table A.11 in Appendix A.3), which is consistent with the association of increased heart rate with deterioration [62]. Additionally, the group showed higher heart rate- ratio amount, indicating a higher tendency of the heart rate to increase rather than decrease over time, and a positive overall slope over time, compared to the negative slope over time in the non-deteriorated group. These findings are supported by Churpek et al, who found that incorporating the slope of heart rate into analysis improved diagnostic accuracy [8].

Lastly, it was notable that from the top 15 features, 9 consisted of complex trend features. This finding suggest that trend features of classical PPG features over time have predictive value, potentially more than the average value of the classical feature alone. This is consistent with the findings that adding trends can be of added value [8, 37].

5.1.3 EMG trend features

A Mann-Whitney U test was also conducted on the EMG features, similar to the PPG feature analysis. It is however important to note that the sample size (38 patients) and the imbalance between the groups (6 patients in the deterioration group) were limiting factors in this analysis.

The results showed that four features were statistically significant, with three of them related to respiratory rate. Surprisingly, the delta and variation features (standard deviation and coefficient of variation) of the respiratory rate in the deteriorating group were lower, contrasting the commonly observed increased variability in sepsis patients, as reported by Bloch et al. [37]. Moreover, the expectation of a higher average respiration rate in the deteriorated group was not met, as the group actually had a lower average respiratory rate. These unexpected results may be due to a bias in the data collection process. The EMG features were primarily collected from periods of good signal quality with limited muscle artifacts. During periods of clinical deterioration, patients may have had higher respiration rates and increased restlessness and movement, which often resulted in worse signal quality and were excluded from the analysis. This may have led to a bias towards measuring respiration rate during times when the patient was relatively stable. Furthermore, three of the six patients who where transferred to the ICU received supplemental oxygen through Optiflow, which could have provided support and had an impact on their breathing effort and respiration frequency.

The results furthermore showed a higher minimal area under the curve, indicating a higher baseline diaphragmatic activity in rest. This aligns with the general knowledge that sepsis causes increased respiratory muscle activity in an effort to ensure adequate oxygenation, leading to higher diaphragmatic activity both at rest and during activity. However, the maximum and mean of the AUC were not higher in the deteriorating group, contrary to expectation. Further investigation is required to gain a better understanding of these unexpected findings.

The statistical analysis did not find significant results in amplitude or breathing intensity features, which may have been due to normalization of the EMG signals. The focus was instead on trend features, such as overall slope and number of trend changes, as changes in respiratory function are considered a sensitive indicator of clinical deterioration [63]. However, there was a lack of high-quality data

for trend analysis due to gaps in the EMG data, caused by patients removing the electrodes or by sweating. The system was not able to produce extended and reliable measurements in patients who where restless or were sweating.

5.1.4 Predictive models

The results suggest that models relying only on vital signs and demographics collected at triage are not highly accurate in predicting deterioration, though they do contain some degree of predictive value. A previous study done in the emergency department with similar data on a larger population (765 patients) found that vital parameters measured at triage had more predictive value. The difference in outcomes between the two studies could be due to the fact that in this study, critically ill patients who where unable to provide informed consent could not be included in the dataset, potentially leading to a dataset with relatively less severe vital parameters. However, it is not verified whether the not included population was in general sicker.

The results of the 55 PPG feature models showed better performance than chance, indicating predictive potential for deterioration. However, overfitting was observed, due to the high number of strongly correlated features. This resulted in models that fit the training data too well, capturing noise instead of underlying patterns, leading to poor performance on the validation data. The KNN model performed better than the LR, SVM and RF classifiers, likely due to its non-parametric nature. Unlike parametric models, KNN does not make assumptions about the distribution of the data, allowing it to capture complex relationships between features and target variables.

The results from using a subset set of PPG features in the model showed improved performance in all classifiers, particularly for the LR, SVM and RF models, where overfitting was reduced. By limiting the number of features, these models became simpler and less likely to capture noise, leading to a better generalization.

The final analysis assessed the impact of incorporating vital signs on model performance. The results indicated a modest improvement in prediction accuracy when vital signs were combined with PPG features. However, the significance of this improvement was not tested. The composite model, with an AUROC score of 0.82 using the KNN model, outperformed all previous models. This result highlights that the PPG trend features contain additional information that is distinct from vital signs, leading to a more accurate prediction.

5.2 Strengths

The study presents several strengths in its method for predicting deterioration, particularly in its ability to detect deterioration in an early stage. There have been studies that showed that individual machine learning models can accurately predict the onset of sepsis ahead of time, including one that used solely PPG signals and deep learning models to perform classification [64, 65]. However, these studies were primarily conducted in ICU settings, where sepsis is often already in an advanced stage. Early detection and intervention are crucial in preventing patients from reaching a more serious condition. The fact that this study was able to achieve successful results in the early stages of deterioration is therefore a notable strength.

Another strength of the study is the use of trends within the PPG signal, which has not been explored previously. Although traditional PPG features can provide information about the cardiovascular status, such as heart rate, stroke volume, systemic vascular resistance, total peripheral resistance, and arterial stiffness [47, 20], the intra-individual changes may provide even more crucial information. Trend analysis allows for the detection of subtle changes over time that may not be evident when considering only the absolute values of PPG features. Previous studies have shown that using trends of vital signs into models significantly improved their accuracy in detecting critical illness in ward settings [8].

Previously, trend features such as minimum, maximum, standard deviation, mean, slope, and delta have been used. However, given that these trend features were based on a limited number of vital sign measurements, the present study has developed more advanced trend features based on continuously available data. These advanced trend features include the ratio of feature increase and decrease, the number of trend changes, the steepest, longest, and overall slopes, and appear to offer even greater insights into the patient's condition.

The robustness of the prediction is another strength of the study, considering that the outcome also includes an increase of two SOFA points, in addition to ICU admission and mortality. While hard outcomes such as ICU admission and mortality tend to be easier to predict, this study used a combination of all three outcomes and achieved a remarkable AUROC score of 0.82. This result indicates that the trend features extracted from the PPG waveforms contain valuable information about organ failure as well.

Finally, the use of PPG technology, which is simple and inexpensive, also adds to the feasibility and practicality of the study. Many patients were willing to wear the wearable and reported no discomfort and the quality of the data collected was sufficient for analysis. In conclusion, the study demonstrated that it is both feasible and practical to use PPG wearables in the prediction of sepsis.

5.3 Limitations

A limitation of this study is the patient population, which consisted of 197 patients, with only 25 of them experiencing deterioration. This imbalanced population may pose challenges for ML models, as they become biased towards the more prevalent class and perform poorly on the less prevalent class. To address this issue, SMOTE was used to balance class distribution, but it is preferable to use real data rather than synthetic data to improve the generalizability of the model to new data.

Another limitation of this study is the method used for pre-processing the EMG data. Template subtraction was used to remove ECG activity, but it was necessary to perform this step twice to remove most of the ECG activity. In some cases, residual ECG activity remained, which resulted in an unreliable signal-to-noise ratio (SNR). As a result, an alternative measure was used to calculate the SNR. In future studies, it may be more effective to use a different approach to more strictly remove ECG activity, such as gating [51]. This method involves the identification of R-peaks, followed by the marking of QRS complexes with a pulse generator having a standardized width. The QRS complexes, along with all other EMG activity, are then removed from the signal at the generated pulses. Although this approach may result in the loss of some EMG signals, it might allow for the use of the SNR and therefore provide a more accurate extraction of good quality EMG data.

As mentioned earlier, the EMG signal is normalized to compare changes over time between patients, instead of absolute values. However, this normalization process means that the absolute effort a patient is putting into their respiration cannot be compared to another patient, but only whether there has been an increase or decrease in effort. Normalization is crucial, but the inability to capture absolute effort is a limitation.

The limitations of the machine learning models used in this study include the absence of a separate test set and the use of the filter method for feature selection. The results are currently based on 5-fold cross-validation, which can indicate performance on unseen data, but still have a risk of overfitting. A separate test set, typically 10% of the data, is usually recommended to reduce this risk. However, this was not possible in this study due to the small size of the dataset. The feature selection method used in this study (filter) does not consider feature interactions or their impact on the model performance. Future research should use the wrapper method instead to select features.

Finally, the plan was to measure both PPG and EMG signals simultaneously to include information from both the cardiovascular and respiratory system in the models. However, due to technical issues, it was not feasible to obtain both types of measurements from a large group of patients. The predictive models were therefore built using only PPG data, but it is expected that combining both will lead to better predictions.

5.4 Future research

To improve the models, it may be useful to include vital signs measured at subsequent time points as well as trend information of the vitals. Incorporating respiration rate, a key predictor of clinical deterioration, could also be valuable [66]. The respiration rate can be estimated from PPG, but accuracy may vary depending on the PPG device and the signal-to-noise ratio [67].

Some of the trend features appear to have the ability to predict clinical deterioration, with the number of trend changes being particularly noteworthy. It may be beneficial to conduct further investigation into this feature, including analyzing additional characteristics such as the minimum, maximum and average intensity of the trend changes. This approach was previously used by Bloch et al, who examined the variability of vital signs using similar features [37]. As previously mentioned, it is important to consider that trend changes may be predictive not only due to the underlying pathophysiology, but also due to the impact of interventions on patients. To better understand the role of these multiple factors, it may be useful to study PPG waveforms in healthy individuals induced with sepsis using Lipopolysaccharide (LPS) without any intervening treatments. LPS is an endotoxin and is recognized as the most potent microbial mediator implicated in the pathogenesis of sepsis and septic shock [68]. Furthermore, the induction of the pathophysiological response with a fixed starting time provides valuable clinical insights.

In addition to examining trends within classic PPG features over a long time period (e.g. hours), it is also of interest to analyze the variability within a feature within a shorter time frame (e.g. 10 minutes). Heart rate variability (HRV) features have been previously studied as potential indicators of clinical deterioration in sepsis patients [69]. A similar analysis can also be performed on PPG features. To do this effectively, it is important to distinguish between PPG waves that deviate from the average wave due to pathophysiological variation and those that may be caused by factors such as movement or reduced contact between the sensor and the finger.

For future research it is interesting if the prediction could be made earlier. Right now, it has been shown that the current model is making a good prediction of deterioration, based on PPG signals from the initial 12 hours. In an ideal scenario, it would be preferable to make this prediction as quickly as possible. Therefore, it is of interest to explore whether the features from the first 6 hours, for example, are also able to provide a reliable prediction.

Lastly, it would be of interest to create patient-specific models for specific groups such as young people and immune-compromised patients. These subgroups may have unique characteristics that could influence their risk of clinical deterioration. For example, young people may have a higher risk of rapid clinical deterioration due to their generally good health and lack of comorbidities. Immune compromised patients may also have an increased risk of clinical deterioration due to their compromised immune systems. By creating predictive models specifically for these subgroups, it may be possible to more accurately identify those at risk of deterioration and provide timely interventions to prevent or mitigate further decline.

5.5 Clinical implications

The early detection of sepsis, a life-threatening condition that can result in septic shock and multiple organ failure, is a critical aspect in patient care. Despite the potential for deterioration within 72 hours, continuous monitoring is not a standard practice in healthcare settings. This study demonstrated that through the analysis of trends from 12-hour PPG measurements using a wearable device, it is possible to predict deterioration after 48 hours. This information provides the nurse with additional insight into the patient's condition, enabling them to closely monitor the patient, initiate antibiotics or fluid therapy, or consult with a physician regarding further management. The initial six hours following the onset of sepsis symptoms are particularly critical for prognosis, and research has shown that appropriate antibiotic treatment within this time frame can reduce the incidence of shock by half [18]. Hence, early identification of patients at risk is of significant importance.

However, a 12-hour measurement may still be considered too long. By refining the algorithm and equipping the system with an alarm feature that triggers an alarm when specific thresholds are exceeded could reduce the wait time for complete measurements. Another approach would be to conduct trend measurements during a shorter duration, such as 6 hours or the typical stay of the patient in the ED to provide earlier predictions.

The ultimate goal of these efforts is to provide early warning to healthcare providers, enabling prompt intervention and reducing in-hospital mortality, shortening hospital stays, and lowering healthcare costs.

6 CONCLUSION

In this study, we analyzed photoplethysmography and diaphragm EMG waveform data to identify trend features that can predict deterioration, including ICU admission, mortality, and a 2-point rise in SOFA score within 48 hours. Using a 12-hour measurement period, we achieved a AUROC score of 0.82. The K-nearest neighbor algorithm was found to be the most effective classification model. Among the photoplethysmography features, the Crest Time-Ratio Amount was most closely related to deterioration, followed by the median stroke volume and the number of trend changes in the diastolic peak amplitude and reflection index and heart rate related features. Unfortunately, no significant trend features were extracted from the EMG data due to gaps in the 12-hour measurement period. Further research is needed to assess the model's generalizability using a separate test set and to determine if predictions can be made earlier than 12 hours.

REFERENCES

[1] L. M. Napolitano, “Sepsis 2018: Definitions and guideline changes,” *Surgical Infections*, vol. 19, pp. 117–125, 2.

[2] K. E. Rudd, S. C. Johnson, K. M. Agesa, K. A. Shackelford, D. Tsoi, D. R. Kievlan, D. V. Colombara, K. S. Ikuta, N. Kissoon, S. Finfer, C. Fleischmann-Struzek, F. R. Machado, K. K. Reinhart, K. Rowan, C. W. Seymour, R. S. Watson, T. E. West, F. Marinho, S. I. Hay, R. Lozano, A. D. Lopez, D. C. Angus, C. J. Murray, and M. Naghavi, “Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the global burden of disease study,” *Lancet (London, England)*, vol. 395, pp. 200–211, 1.

[3] M. Singer, C. S. Deutschman, C. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. R. Bernard, J. D. Chiche, C. M. Coopersmith, R. S. Hotchkiss, M. M. Levy, J. C. Marshall, G. S. Martin, S. M. Opal, G. D. Rubenfeld, T. D. Poll, J. L. Vincent, and D. C. Angus, “The third international consensus definitions for sepsis and septic shock (sepsis-3),” *JAMA*, vol. 315, p. 801, 2.

[4] S. W. Glickman, C. B. Cairns, R. M. Otero, C. W. Woods, E. L. Tsalik, R. J. Langley, J. C. V. Velkinburgh, L. P. Park, L. T. Glickman, V. G. Fowler, S. F. Kingsmore, and E. P. Rivers, “Disease progression in hemodynamically stable patients presenting to the emergency department with sepsis,” *Academic emergency medicine : official journal of the Society for Academic Emergency Medicine*, vol. 17, pp. 383–390.

[5] F. Lellouche and E. L’her, “Usual and advanced monitoring in patients receiving oxygen therapy,” *Respiratory care*, vol. 65, pp. 1591–1600, 10.

[6] J. P. Leenen, H. J. Rasing, J. D. Van Dijk, C. J. Kalkman, L. Schoonhoven, and G. A. Patijn, “Feasibility of wireless continuous monitoring of vital signs without using alarms on a general surgical ward: A mixed methods study,” *PloS one*, no. 3, mar.

[7] J. D. Young, “The heart and circulation in severe sepsis,” *British Journal of Anaesthesia*, vol. 93, pp. 114–120, 7 2004.

[8] M. M. Churpek, R. Adhikari, and D. P. Edelson, “The value of vital sign trends for detecting clinical deterioration on the wards,” *Resuscitation*, vol. 102, p. 1, 5.

[9] I. J. Brekke, L. H. Puntervoll, P. B. Pedersen, J. Kellett, and M. Brabrand, “The value of vital sign trends in predicting and monitoring clinical deterioration: A systematic review,” *PloS one*, no. 1, jan.

[10] J. E. Gotts and M. A. Matthay, “Sepsis: pathophysiology and clinical management,” *BMJ*, may.

[11] B. Gyawali, K. Ramakrishna, and A. S. Dhamoon, “Sepsis: The evolution in definition, pathophysiology, and management,” *SAGE Open Medicine*, vol. 7, p. 205031211983504, 1.

[12] A. E. Jones, S. Trzeciak, and J. A. Kline, “The sequential organ failure assessment score for predicting outcome in patients with severe sepsis and evidence of hypoperfusion at the time of emergency department presentation,” *Critical care medicine*, vol. 37, p. 1649.

[13] C. T. Esmon, "The interactions between inflammation and coagulation," *British journal of haematology*, vol. 131, pp. 417–430, 11.

[14] I. Eckle, R. Seitz, R. Egbring, G. Kolb, and K. Havemann, "Protein C degradation in vitro by neutrophil elastase," *Biological chemistry Hoppe-Seyler*, no. 11, pp. 1007–1014.

[15] R. S. Hotchkiss, K. W. Tinsley, P. E. Swanson, R. E. Schmieg, J. J. Hui, K. C. Chang, D. F. Osborne, B. D. Freeman, J. P. Cobb, T. G. Buchman, and I. E. Karl, "Sepsis-induced apoptosis causes progressive profound depletion of B and CD4+ T lymphocytes in humans," *Journal of immunology (Baltimore, Md. : 1950)*, no. 11, pp. 6952–6963, jun.

[16] B. J. Biemond, M. Levi, H. T. Cate, T. V. der Poll, H. R. Buller, C. E. Hack, and J. W. T. Cate, "Plasminogen activator and plasminogen activator inhibitor i release during experimental endotoxaemia in chimpanzees: effect of interventions in the cytokine and coagulation cascades," *Clinical science (London, England : 1979)*, vol. 88, pp. 587–594.

[17] W. Heagy, K. Nieman, C. Hansen, M. Cohen, D. Danielson, and M. A. West, "Lower levels of whole blood LPS-stimulated cytokine release are associated with poorer clinical outcomes in surgical ICU patients," *Surgical infections*, no. 2, pp. 171–180.

[18] G. Polat, R. A. Ugan, E. Cadirci, and Z. Halici, "Sepsis and Septic Shock: Current Treatment Strategies and New Approaches," *The Eurasian Journal of Medicine*, no. 1, p. 53, feb.

[19] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological measurement*, vol. 28, 3 2007. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/17322588/

[20] M. Elgendi, "On the Analysis of Fingertip Photoplethysmogram Signals," *Current Cardiology Reviews*, no. 1, p. 14, jun.

[21] T. Tamura, "Current progress of photoplethysmography and SPO2 for health monitoring," *Biomedical Engineering Letters*, no. 1, p. 21, feb.

[22] A. Dzedzickis, A. Kaklauskas, and V. Bucinskas, "Human emotion recognition: Review of sensors and methods," *Sensors (Switzerland)*, vol. 20, 2 2020.

[23] W. Murray and P. A. Foster, "The peripheral pulse wave: information overlooked," *Journal of clinical monitoring*, no. 5, pp. 365–377.

[24] E. C. P. Chua, S. J. Redmond, G. McDarby, and C. Heneghan, "Towards using photoplethysmogram amplitude to measure blood pressure during sleep," *Annals of biomedical engineering*, no. 3, pp. 945–954, mar.

[25] A. A. Awad, A. S. Haddadin, H. Tantawy, T. M. Badr, R. G. Stout, D. G. Silverman, and K. H. Shelley, "The relationship between the photoplethysmographic waveform and systemic vascular resistance," *Journal of clinical monitoring and computing*, no. 6, pp. 365–372, dec.

[26] W. M. Jubadi and S. F. A. M. Sahak, "Heartbeat monitoring alert via SMS," *2009 IEEE Symposium on Industrial Electronics and Applications, ISIEA 2009 - Proceedings*, vol. 1, no. Isiea, pp. 1–5, 2009.

[27] S. R. Alty, N. Angarita-Jaimes, S. C. Millasseau, and P. J. Chowienczyk, "Predicting arterial stiffness from the digital volume pulse waveform," *IEEE transactions on bio-medical engineering*, no. 12, pp. 2268–2275, dec.

[28] J. Mok Ahn, "New Aging Index Using Signal Features of Both Photoplethysmograms and Acceleration Plethysmograms," *Healthcare Informatics Research*, no. 1, p. 53, jan.

[29] C. Chamos, L. Vele, M. Hamilton, and M. Cecconi, "Less invasive methods of advanced hemodynamic monitoring: principles, devices, and their role in the perioperative hemodynamic optimization," *Perioperative Medicine*, no. 1, dec.

[30] L. Wang, E. Pickwell-MacPherson, Y. P. Liang, and Y. T. Zhang, "Noninvasive cardiac output estimation using a novel photoplethysmogram index," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, pp. 1746–1749.

[31] K. Takazawa, N. Tanaka, M. Fujita, O. Matsuoka, T. Saiki, M. Aikawa, S. Tamura, and C. Ibukiyama, "Assessment of vasoactive agents and vascular aging by the second derivative of photoplethysmogram waveform," *Hypertension (Dallas, Tex. : 1979)*, no. 2, pp. 365–370.

[32] I. Imanaga, H. Hara, S. Koyanagi, and K. Tanaka, "Correlation between wave components of the second derivative of plethysmogram and arterial distensibility," *Japanese heart journal*, no. 6, pp. 775–784.

[33] M. Elgendi, I. Norton, M. Brearley, D. Abbott, and D. Schuurmans, "Detection of a and b waves in the acceleration photoplethysmogram," *BioMedical Engineering Online*, no. 1, pp. 1–18, sep.

[34] P. R. Patel and S. Bechmann, "Elevated hemidiaphragm," *Radiopaedia.org*, 8.

[35] E. Z. Hawkes, A. V. Nowicky, and A. K. McConnell, "Diaphragm and intercostal surface emg and muscle performance after acute inspiratory muscle loading," *Respiratory Physiology Neurobiology*, vol. 155, pp. 213–219, 3 2007.

[36] J. Graßhoff, E. Petersen, F. Farquharson, M. Kustermann, H. J. Kabitz, P. Rostalski, and S. Walterspacher, "Surface EMG-based quantification of inspiratory effort: a quantitative comparison with P es," *Critical Care*, no. 1, pp. 1–12, dec.

[37] E. Bloch, T. Rotem, J. Cohen, P. Singer, and Y. Aperstein, "Machine Learning Models for Analysis of Vital Signs Dynamics: A Case for Sepsis Onset Prediction," *Journal of Healthcare Engineering*, vol. 2019, 2019.

[38] Loukas Serafeim, "What is Machine Learning: Supervised, Unsupervised, Semi-Supervised and Reinforcement learning methods," 2020. [Online]. Available: https://towardsdatascience.com/what-is-machine-learning-a-short-note-on-supervised-unsupervised-semi-supervised-and-aed

[39] H. Belyadi and A. Haghighat, "Supervised learning," *Machine Learning Guide for Oil and Gas Using Python*, pp. 169–295, jan 2021.

[40] "Support Vector Machine (SVM) Algorithm." [Online]. Available: https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm

[41] L. Hao, J. Kim, S. Kwon, and I. D. Ha, "Deep Learning-Based Survival Analysis for High-Dimensional Survival Data," *Mathematics 2021, Vol. 9, Page 1244*, no. 11, p. 1244, may.

[42] "Confusion Matrix ." [Online]. Available: https://rapidminer.com/glossary/confusion-matrix/

[43] E. Ter Avest, B. C. Van Munster, R. J. Van Wijk, S. Tent, S. Ter Horst, T. T. Hu, L. E. Van Heijst, F. S. Van Der Veer, F. E. Van Beuningen, J. C. Ter Maaten, and H. R. Bouma, "Cohort profile: Cohort profile of Acutelines: a large data/biobank of acute and emergency medicine," *BMJ Open*, no. 7, jul.

[44] P. Tanabe, R. Gimbel, P. R. Yarnold, and J. G. Adams, "The Emergency Severity Index (version 3) 5-level triage system scores predict ED resource consumption," *Journal of Emergency Nursing*, no. 1, pp. 22–29.

[45] M. McCaffery, O. Onikoyi, D. Rodrigopulle, A. Syed, S. Jones, L. Mansfield, and M. G Krishna, "Sepsis-review of screening for sepsis by nursing, nurse driven sepsis protocols and development of sepsis hospital policy/protocols," *Nursing and Palliative Care*, vol. 1, no. 2, pp. 33–37, 2016.

[46] Y. Li, C. Yan, Z. Gan, X. Xi, Z. Tan, J. Li, and G. Li, "Prognostic values of SOFA score, qSOFA score, and LODS score for patients with sepsis." *Annals of palliative medicine*, vol. 9, no. 3, pp. 1037–1044, may 2020.

[47] J. Park, H. S. Seok, S. S. Kim, and H. Shin, "Photoplethysmogram Analysis and Applications: An Integrative Review," *Frontiers in Physiology*, vol. 12, p. 2511, mar 2022.

[48] "Patiënten met slaapapneu sneller behandeld dankzij OSAsense." [Online]. Available: https://www.utwente.nl/nieuws/2019/11/42540/patienten-met-slaapapneu-sneller-behandeld-dankzij-osasense

[49] Q. Li and G. D. Clifford, "Dynamic time warping and machine learning for signal quality assessment of pulsatile signals," *Physiological measurement*, no. 9, pp. 1491–1501.

[50] E. Petersen, J. Sauer, J. Grabhoff, and P. Rostalski, "Removing Cardiac Artifacts from Single-Channel Respiratory Electromyograms," *IEEE Access*, vol. 8, pp. 30 905–30 917, 2020.

[51] R. W. van Leuteren, G. J. Hutten, C. G. de Waal, P. Dixon, A. H. van Kaam, and F. H. de Jongh, "Processing transcutaneous electromyography measurements of respiratory muscles, a review of analysis techniques," *Journal of electromyography and kinesiology : official journal of the International Society of Electrophysiological Kinesiology*, pp. 176–186, oct.

[52] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, 1985.

[53] A. Fratini, M. Cesarelli, P. Bifulco, and M. Romano, "Relevance of motion artifact in electromyography recordings during vibration treatment," *Journal of Electromyography and Kinesiology*, vol. 19, no. 4, pp. 710–718, aug 2009.

[54] "Prominence - MATLAB Simulink - MathWorks Benelux." [Online]. Available: https://nl.mathworks.com/help/signal/ug/prominence.html

[55] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society: Series B (Methodological)*, no. 1, pp. 289–300, jan.

[56] A. Sapra, A. Malik, and P. Bhandari, "Vital Sign Assessment," *StatPearls*, may.

[57] S. Jain and L. M. Iverson, "Glasgow Coma Scale," no. 5, pp. 1–5, jun.

[58] Z. Hu, G. B. Melton, E. G. Arsoniadis, Y. Wang, M. R. Kwaan, and G. J. Simon, "Strategies for handling missing clinical data for automated surgical site infection detection from the electronic health record," *Journal of biomedical informatics*, p. 112, apr.

[59] Z. Xu, D. Shen, Y. Kou, and T. Nie, "A Synthetic Minority Oversampling Technique Based on Gaussian Mixture Model Filtering for Imbalanced Data Classification," *IEEE transactions on neural networks and learning systems*.

[60] F. J. Dankers, A. Traverso, L. Wee, and S. M. van Kuijk, "Fundamentals of Clinical Data Science ," in *Fundamentals of Clinical Data Science*. Springer, dec 2018, ch. Prediction Modeling…, pp. 101–120.

[61] Y. Kakihana, T. Ito, M. Nakahara, K. Yamaguchi, and T. Yasuda, "Sepsis-induced myocardial dysfunction: pathophysiology and management," *Journal of Intensive Care*, no. 1.

[62] J. E. Stevenson, J. Israelsson, G. C. Nilsson, G. I. Petersson, and P. A. Bath, “Recording signs of deterioration in acute patients: The documentation of vital signs within electronic health records in patients who suffered in-hospital cardiac arrest,” *Health Informatics Journal*, vol. 22, no. 1, pp. 21–33, 2016.

[63] H. Cahill, A. Jones, R. Herkes, K. Cook, A. Stirling, T. Halbert, A. Yates, S. Lal, A. Gardo, R. Donnelly, and D. J. Gattas, “Introduction of a new observation chart and education programme is associated with higher rates of vital-sign ascertainment in hospital wards,” *BMJ Quality Safety*, no. 9, pp. 791–796, sep.

[64] L. M. Fleuren, T. L. Klausch, C. L. Zwager, L. J. Schoonmade, T. Guo, L. F. Roggeveen, E. L. Swart, A. R. Girbes, P. Thoral, A. Ercole, M. Hoogendoorn, and P. W. Elbers, “Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy,” *Intensive care medicine*, vol. 46, pp. 383–400, 3 2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/31965266/

[65] S. Lombardi, P. Partanen, P. Francia, I. Calamai, R. Deodati, M. Luchini, R. Spina, and L. Bocchi, “Classifying sepsis from photoplethysmography,” *Health Information Science and Systems*, vol. 10, no. 1, pp. 1–12, dec 2022. [Online]. Available: https://link.springer.com/article/10.1007/s13755-022-00199-3

[66] K. Mochizuki, R. Shintani, K. Mori, T. Sato, O. Sakaguchi, K. Takeshige, K. Nitta, and H. Imamura, “Importance of respiratory rate for the prediction of clinical deterioration after emergency department discharge: a single□center, case–control study,” *Acute Medicine Surgery*, no. 2, p. 172, apr.

[67] T. Iqbal, A. Elahi, S. Ganly, W. Wijns, and A. Shahzad, “Photoplethysmography-Based Respiratory Rate Estimation Algorithm for Health Monitoring Applications,” *Journal of Medical and Biological Engineering*, no. 2, p. 242, apr.

[68] S. M. Opal, “Endotoxins and other sepsis triggers,” *Contributions to nephrology*, pp. 14–24, jun.

[69] D. P. Barnaby, S. M. Fernando, C. L. Herry, N. B. Scales, E. J. Gallagher, and A. J. Seely, “Heart Rate Variability, Clinical and Laboratory Measures to Predict Future Deterioration in Patients Presenting With Sepsis,” *Shock (Augusta, Ga.)*, no. 4, pp. 416–422, apr.

A APPENDIX

A.1 Cardiac Artifact Removal

In this section, the process of the initial removal of cardiac activity is described in more detail. To accomplish this task, the template subtraction method, as described in [50] was used. First, the R-peaks were detected using the Pan-Tompkins method [52]. The algorithm was designed to identify the QRS complexes in the electrocardiography (ECG) signal, therefore, a 2nd order Butterworth band-pass filter with a frequency range of 5-15 Hz was used to eliminate muscle noise.

Next, the the derivative of the ECG signal was computed and the signal was squared. A window of ten seconds was used to inspect the signal for extreme artifacts, defined as an amplitude deviation higher than 3000 μV within the 10 second window. If the criterion was fulfilled, the window was discarded. The start and end of a QRS complex were determined as the valley and peak of the moving-window integration (MWI) of the ECG signal. The position of the R-peak was defined as the minimum value of the ECG value between the start and end of the QRS complex, as demonstrated in Figure A.1. The MWI was calculated using a window length of 200 ms.

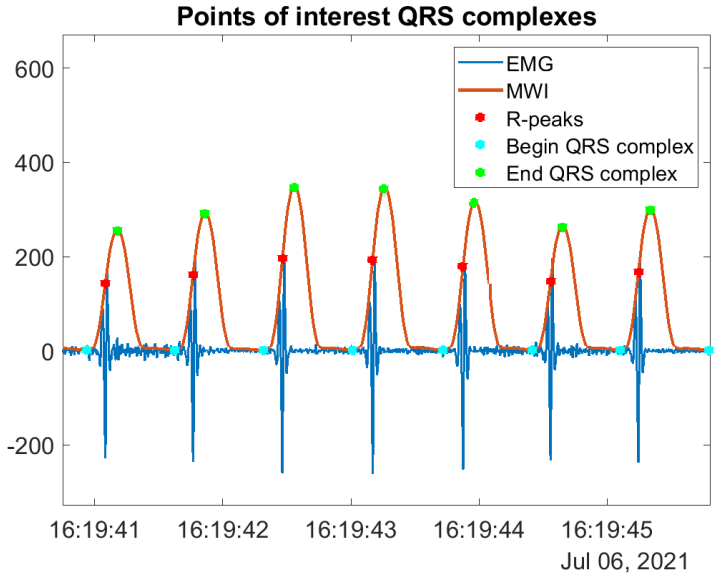


Figure A.1: Points of interest of the QRS complexes

To remove the ECG artifact from the EMG signal, a subtraction template was generated for each ECG beat. The process of generating the template involved averaging 21 ECG beats. This average included the current beat, ten beats before it, and ten beats after it, as shown in the top graph of Figure A.2. If the current beat was one of the first ten or last ten beats, the average was taken using all available beats before or after it, respectively, resulting in fewer than 21 beats being used. The beats were then aligned based on the R-peaks, and the length of the template window was set to match the length of the current beat.

To align the template and the current ECG beat, their correlation was maximized, as demonstrated in the middle graph of Figure A.2. The offset and amplitude of the template were adjusted using linear regression. Finally, the template was subtracted from the EMG signal, resulting in last graph of Figure A.2.

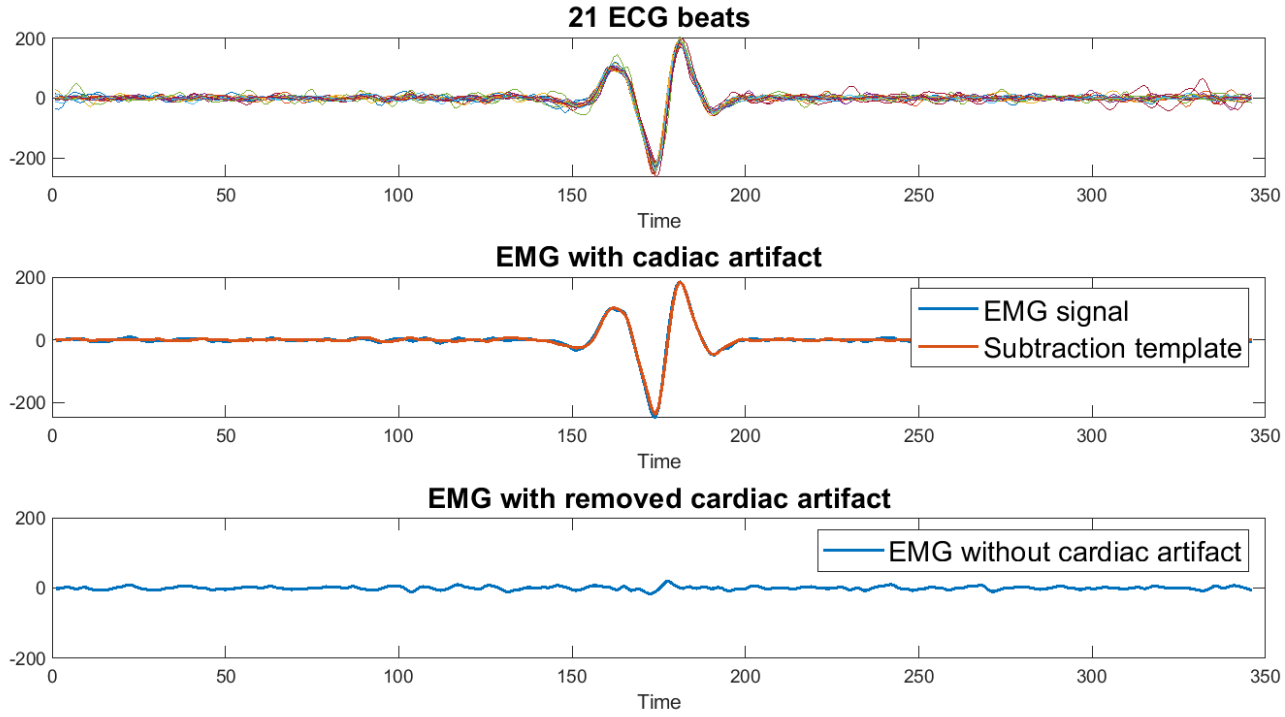
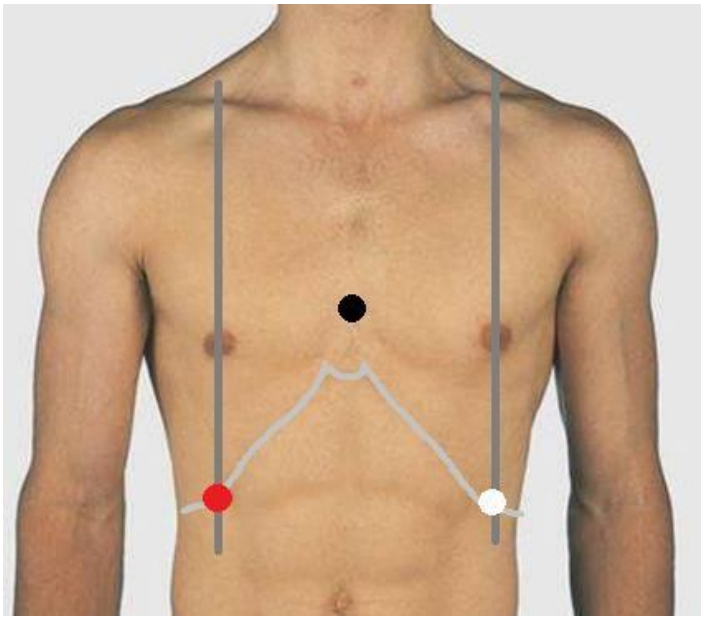


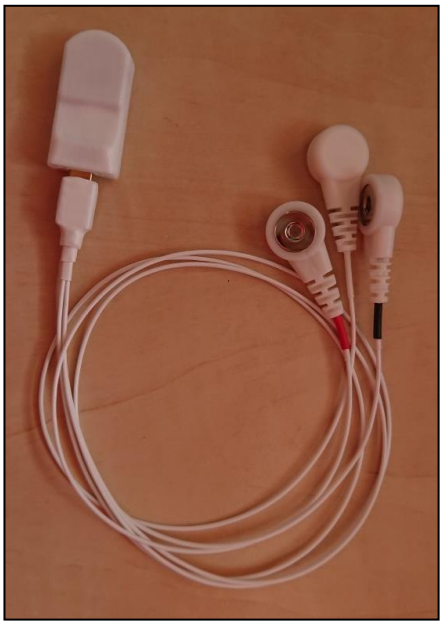
Figure A.2: Subtraction template

A.2 Protocols wearables

Handleiding

METINGEN MET ITOM

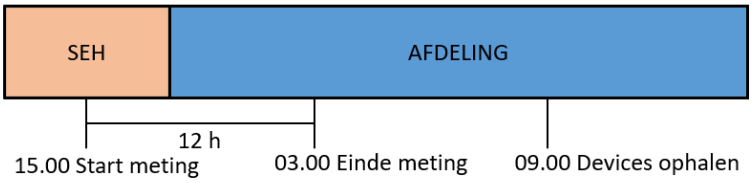




1. Werkwijze

1.1. Meetmomenten

Er wordt 1 meting verricht van 1 x 12 uur. Deze meting wordt op de SEH gestart en loopt door op de afdeling waar de patiënt daarna wordt heengebracht. Het apparaat stopt de meting vanzelf, doordat de batterij leeg is of de opslag vol is. De zaaldienst of avonddienst moet de sensor 12 uur later er weer afhalen en op de SEH aan de stroom leggen. Hieronder een voorbeeld van hoe dit in de praktijk werkt.

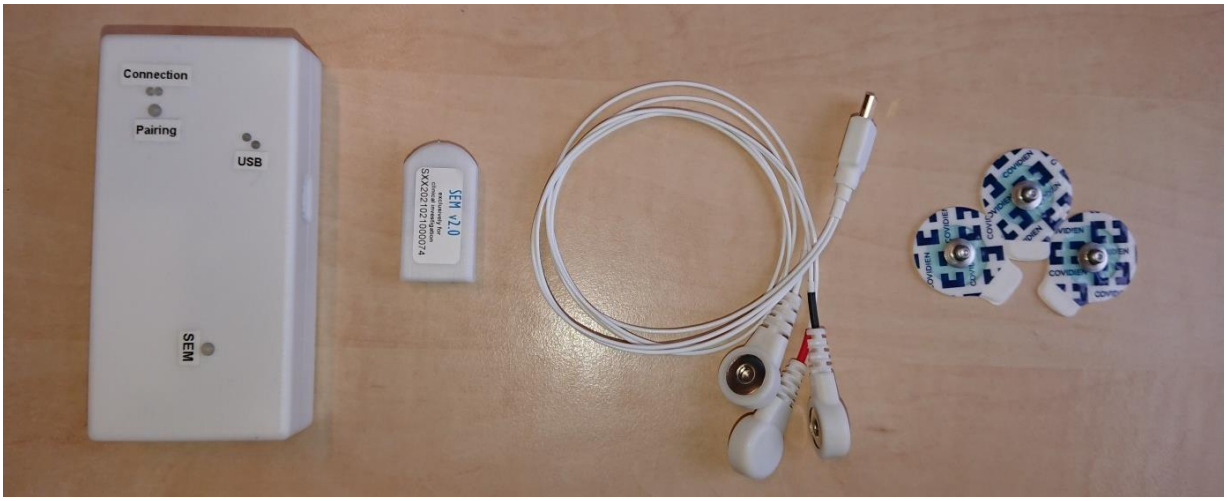


1.2. Benodigdheden

Voor het aansluiten van de ItoM zijn de volgende onderdelen nodig:

- UMCG Laptop: '16-000-7381'
- Laptop HP oplader
- Docking station
- USB naar micro USB kabel (zwarte kabel)
- Sensor module (SEM)
- Elektrode kabels
- Elektrodes
- Alcohol
- Doekjes
- Tape
- Flyer voor de verpleegkundige
- Oranje polsbandje speciaal voor wearables

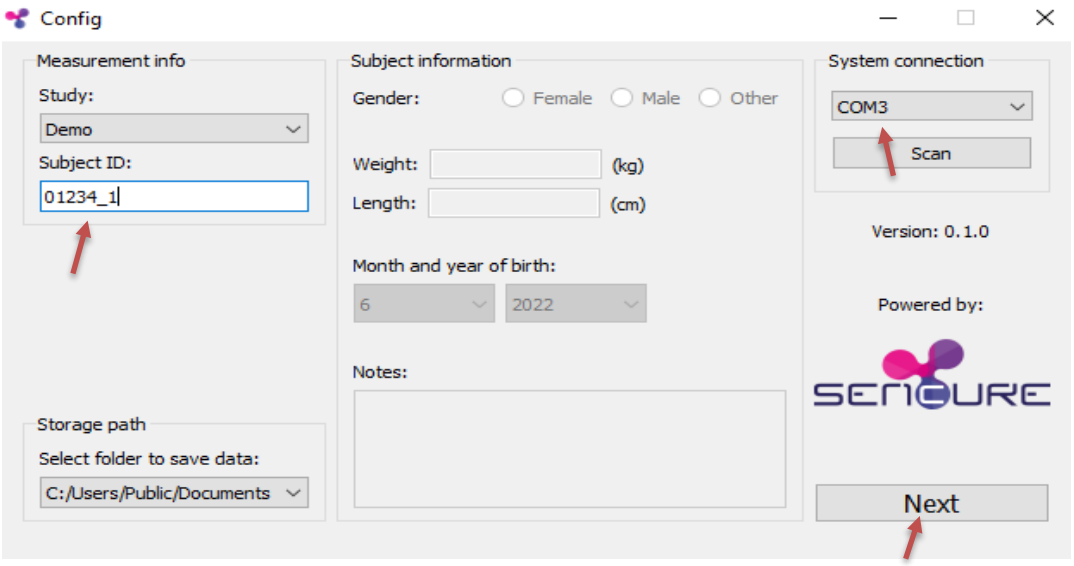
Bijna alle benodigdheden liggen in de staande researchkast. Alcohol en gaasjes kan je bij/in de patiëntkamer vinden.



Figuur 1 – vlnr: Docking station, SEM, elektrode kabels, elektrodes

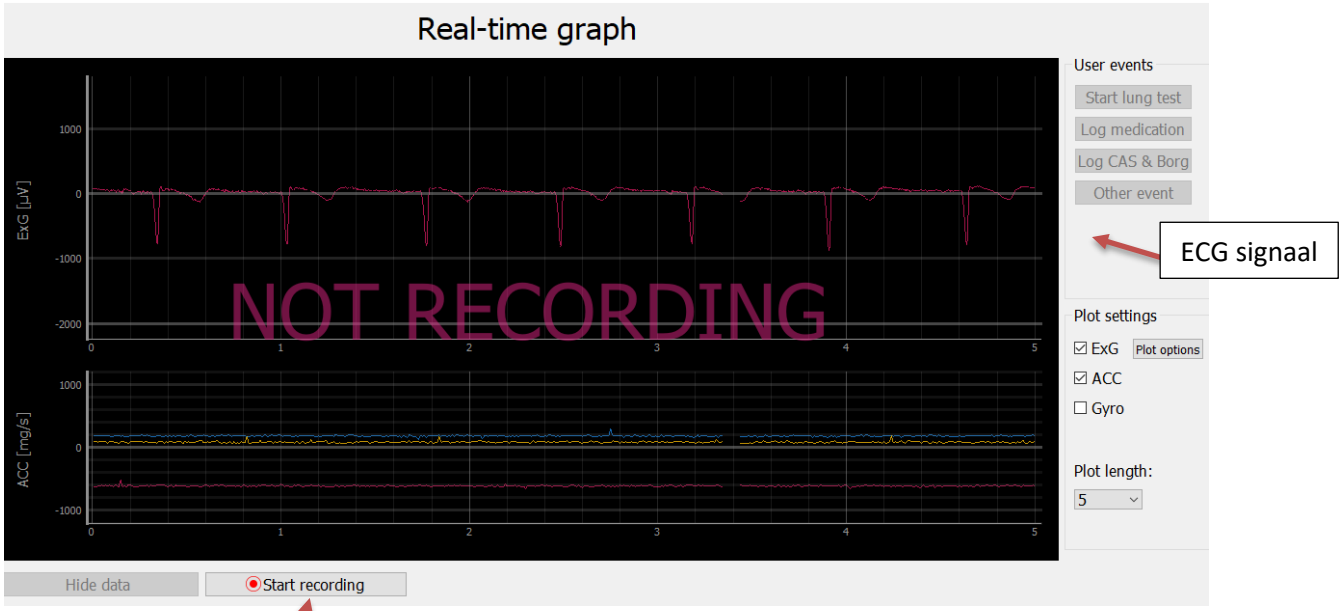
1.3. ItoM meting: SEM instellen en aansluiten

1. Start de laptop op en log in
2. De ItoM applicatie start automatisch op (duurt wel eventjes).
3. Verbind de docking station met de laptop met de zwarte USB-kabel en druk de SEM in de dockingstation. **Let op dat deze met de acutelines sticker bovenop er wordt ingedrukt.**
4. Controleer de batterijstatus van de SEM (led 'SEM')
    a. Groene LED: opgeladen
    b. Gele/oranje/rode led: aan het opladen
5. Controleer de koppelingsstatus van de SEM (led 'Pairing')
    a. Groene led: correct gekoppeld
    b. Geen groene led: voer de SEM opnieuw in in de docking station
6. Voer de juiste settings in het 'Config' scherm →
    a. Voer bij 'Study' → 'Demo' in
    b. Voer bij Subject ID het juiste deelnemer ID + bezoeknummer in (01234_01)
        - Subject ID= 1234 en bezoeknummer = 1
    c. Klik bij 'System connection' → COM 3 in en klik op 'next'. Als er geen COM poort geselecteerd kan worden, doe dan de usb kabel er opnieuw in.



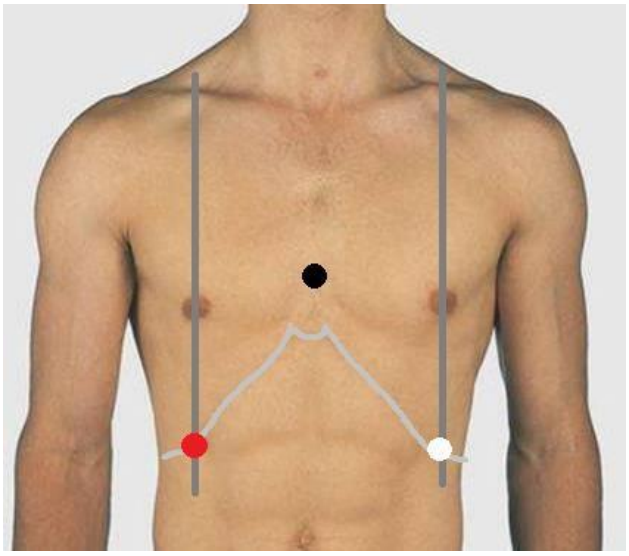
7. Haal de SEM uit de docking station en verbindt deze met de elektrodekabels. Reinig met alcohol. Als het goed is wordt het lampje van de SEM eerst groen en gaat dan blauw/donkerblauw knipperen. Dit betekent dat hij verbinding heeft.
8. Verbindt de elektrodekabels alvast met de electroden voordat ze op de patiënt geplakt worden.

9. Neem nu de laptop inclusief oplader, docking station, SEM met elektrodekabels en elektrodes en tape mee naar de patiënt.
10. Plak de elektrodes op de patiënt. Bij een behaarde borst, eerst plaatselijk scheren. De scheermesjes liggen bij de SEM + elektrodes in de houten bak.
    Zorg ervoor dat je de juiste elektrodekabel met de juiste elektrode verbindt!
    a. Zwart (referentie): op 3e intercostaal ruimte, midsternaal (zie afbeelding)
    b. Rood (rechts) en wit (links): op de ribbenboog, midclaviculair (zie afbeelding)
11. Maak de SEM vast met een stukje tape aan de borstkas.
12. Check nu of er een hartslag signaal wordt geplot in de applicatie.



13. Klik nu op 'Start Recording'.
14. Als het niet lukt om signaal te krijgen, bel dan Anna Schoonhoven!
15. Zet de laptop op een veilig plek en zorg dat de laptop opgeladen wordt.
    a. **Om te zorgen de dat meting doorgaat als de patiënt naar de afdeling gaat moet de RA met de laptop mee lopen als de patiënt opgenomen wordt.**
    b. **Houdt dus in de gaten wanneer de patiënt opgenomen wordt. Dit kan je in EPIC zien in het SEH-trackboard bij vervolg. Of je bespreekt dit met de verpleegkundige en vraagt of hij/zij je komt ophalen (houdt nog steeds wel zelf in de gaten!)**
    c. **Zorg ervoor dat de laptop weer opgeladen wordt op de afdeling en dat hij niet dichtgeklapt wordt!**

16. Kruis op de flyer voor de verpleegkundige aan welke wearables er bij de patiënt worden aangesloten.
17. Vul de 'research notitie' in op Epic.
18. Vul in op Teams dat de ItoM is aangesloten en wanneer deze afgehaald moet worden door de zaaldienst of avonddienst.
19. Vul de starttijd van de meting van ItoM in Redcap.



1.4 Verpleegkundige informeren

20. Plak de flyer voor de verpleegkundige met wat tape aan het voeteneinde van het bed, zodat deze goed zichtbaar is voor alle verpleegkundigen.
21. **Informeer de verpleegkundige dat de patiënt aangesloten is op de wearables**. En vraag of hij/zij je wilt waarschuwen als de patiënt naar de afdeling gaat
22. **Loop met de patiënt en verpleegkundige mee naar de afdeling als de patiënt opgenomen wordt. Informeer de verpleegkundige van de afdeling mondeling over de metingen met de wearables en wijs hen op de flyer voor als ze meer informatie willen.**

1.4. Metingen loskoppelen (ZAALDIENST OF AVONDDIENST)

Kijk in RedCap op het Productie Dashboard 'Wearable Dashboard' en op Teams of er een meting bezig is en hoelaat de meting is gestart. Als er 12 uur voorbij zijn gedaan, dan moeten de ItoM losgekoppeld worden. Vul de eindtijd van de meting in op Redcap en op Teams.

ItoM loskoppelen en schoonmaken

1. Druk op 'Stop Recording' in de ItoM applicatie op de laptop.
2. Sluit de applicatie af.
3. Koppel de elektrodekabel los van de elektrodes.
4. Haal voorzichtig de SEM los van de huid.
5. Verwijder voorzichtig de elektrodes van de huid.
6. Maak de SEM en de elektrodekabel schoon met een gaasje met 70% alcohol. **Let goed op dat de SEM niet kwijtraakt tussen de dekens!**
7. Neem **alles** mee terug naar de research kamer. Alleen de elektrodeplakkers mogen weggegooid worden.
8. Plaats de laptop in de oplader.
9. Laad de SEM op door hem in de docking station te stoppen en te verbinden met de laptop of met een computer.

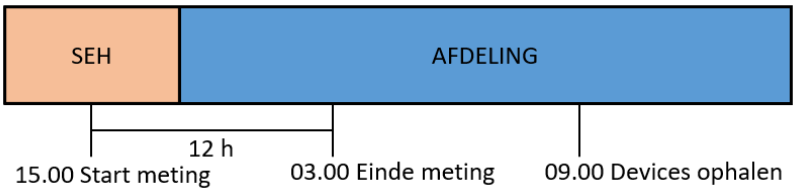
Handleiding

METINGEN MET OSASENSE

1. Werkwijze

1.1. Meetmomenten

Er wordt 1 meting verricht van 1 x 12 uur. Deze meting wordt op de SEH gestart en loopt door op de afdeling waar de patiënt daarna wordt heengebracht. Het apparaat stopt de meting vanzelf, doordat de batterij leeg is of de opslag vol is. De zaaldienst of avonddienst moet de sensor 12 uur later er weer afhalen en op de SEH aan de stroom leggen. Hieronder een voorbeeld van hoe dit in de praktijk werkt.



1.2. Benodigdheden

Voor het aansluiten van het OSAsense meetsysteem zijn de volgende onderdelen nodig:

- UMCG Laptop: '16-000-7377'
- Laptop oplader
- OSAsense meetsysteem (zit in een zwarte bewaarhoes)
- Vingertipsensor
- Oplaad/verbindingskabel
- Flyer voor de verpleegkundige
- Oranje polsbandje speciaal voor wearables

Bijna alle benodigdheden liggen in de staande researchkast. Alcohol en gaasjes kan je bij/in de patiëntkamer vinden.

1.3. OSAsense instellen

1. Pak de UMCG laptop, de OSAsense en een vingertipsensor uit de kast
2. Start de UMCG laptop op
3. Log in
    a. De pop-up van *ED RESEARCH* kan je wegklikken
    b. Er komt ook een pop-up van Citrix in beeld, deze kan je wegklikken, maar hij blijft terugkomen. Je kunt de pop-up ook opzij slepen en verder negeren.
4. Open de OSAsense applicatie
5. Klik op "OSAsense S18 Instellen"
6. Verbind de OSAsense met de computer met de witte kabel
    a. Als het goed is komt de OSAsense bij stap 1 in het scherm te staan.
    b. Gebeurt dit niet: koppel de OSAsense los en opnieuw weer aan of beweeg de kabel voorzichtig totdat de OSAsense wel herkend wordt.
7. Controleer in "Redcap->productie" in het report "OSAsense" welke patiënt codes al gebruikt zijn. Deze lijst vind je in het menu links.

Deelnemer ID subject_id	Event Name redcap_event_name	Repeat Instance redcap_repeat_instance	Screen ID screen_id	OSAsense ID osasense_id
1646	Visit	1	4928	abb
415	Visit	2	4599	aba
1570	Visit	1	4716	aaz
1606	Visit	1	4816	aay
1418	Visit	1	4256	aax
1364	Visit	1	4115	AAW
1288	Visit	1	3916	aav
1321	Visit	1	4000	AAU
1455	Visit	1	4349	aas
1380	Visit	1	4160	aar
1268	Visit	1	3870	

8. Koppel de patiënt in “Redcap->Productie->koppeling” aan een ongebruikte OSAsense code.
9. Vul de begintijd van de meting in in Redcap.
10. Selecteer die patiënt code in de OSAsense app.
    a. **Als er geen/weinig ongebruikte patiënt code zijn.**
11. Klikt op “OSAsense S18 Instellen voor: [patiënt code]”.
12. Koppel de OSAsense los en sluit de applicatie af. De OSAsense is nu klaar voor gebruik.

1.4. OSAsense instellen

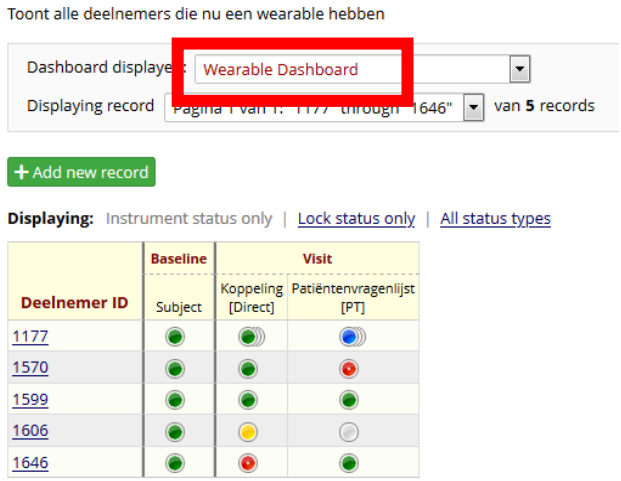
13. Sluit de sensor aan op een vinger van de patiënt op de aangegeven methode
    a. Op de pleister van de sensor staat aangegeven hoe deze om de vinger van de patiënt geplakt moet worden.
14. Doe de OSAsense om de pols van de patiënt en sluit de sensor aan op de OSAsense.
    a. Als het goed is gaat de rode led van de elektrode branden, knippert het rondje linksboven in het schermpje en komt na een paar seconden de saturatie en hartslag in het scherm. De meting is dan begonnen.
    b. Zolang de sensor verbonden is kan de OSAsense niet uitgezet worden. Het scherm gaat vanzelf uit, maar de meting blijft doorgaan

Verpleegkundige informeren

15. Plak de flyer voor de verpleegkundige met wat tape aan het voeteneinde van het bed, zodat deze goed zichtbaar is voor alle verpleegkundigen.
16. **Informeer de verpleegkundige dat de patiënt aangesloten is op de wearables**. En vraag of hij/zij je wilt waarschuwen als de patiënt naar de afdeling gaat.

1.5 Metingen loskoppelen

Kijk in RedCap op het Productie Dashboard ‘Wearable Dashboard’ of er een meting bezig is en hoe laat de meting is gestart. Als er 12 uur voorbij zijn gedaan, dan moeten de ItoM en OSAsense losgekoppeld worden. Vul de eindtijd van de meting in in “Redcap->Productie->koppeling”.



OSAsense loskoppelen en schoonmaken

1. Haal de sensor van de vinger
2. Maak het polshorloge los
3. Koppel de sensor los van het horloge
4. Gooi de sensor weg
5. **Ontsmet het horloge met een gaasje met 70% alcohol**

1.6 Metingen uitlezen en uploaden

1. Start de UMCG laptop op
2. Log in.
    a. De pop-up van *ED RESEARCH* kan je wegklikken
    b. Er komt ook een pop-up van Citrix in beeld, deze kan je wegklikken, maar hij blijft terugkomen. Je kunt de pup-up ook opzij slepen en verder negeren.
3. Open de OSAsense applicatie
4. Klik op “OSAsense S18 Uitlezen”
5. Verbind de OSAsense met de computer, Als het goed is komt de OSAsense bij stap 1 in het scherm te staan. Gebeurt dit niet: koppel de OSAsense los en opnieuw weer aan of beweeg de kabelvoorzichtig totdat de OSAsense wel herkend wordt.
6. Klikt op “Uitlezen & Uploaden”.
7. Op de OSAsense verschijnt een laadbalk. Het uitlezen kan enkele minuten duren.
8. Als het uploaden gelukt is, verschijnt de melding “De data van de OSAsense S18 is uitlezen, geupload en geanalyseerd.” in de applicatie.
9. Sluit de applicatie af
10. Laat de OSAsense verbonden met de laptop om deze op te laden voor de volgende meting.

A.3 Tables PPG features

Table A.1: Trend features of the systolic peak amplitude for deteriorated and non-deteriorated group.

		Deteriorated (n=24)	**Non-deteriorated (n=165)**	**p-value**
Feature	**Trend**			
Systolic Peak Amplitude	Minimum	-	-	-
	Maximum	-	-	-
	Delta	-	-	-
	Mean	-	-	-
	Median	-	-	-
	Standard deviation	0.40 [0.39 - 0.42]	0.40 [0.39 - 0.41]	0.33
	Coefficient of Variation	0.91 [0.79 - 1.02]	0.87 [0.79 - 0.93]	0.13
	Number of trend changes	0.05 [0.04 - 0.06]	0.05 [0.04 - 0.05]	0.06
	Ratio time	1.01 [0.85 - 1.18]	0.93 [0.80 - 1.06]	0.17
	Ratio amount	1.02 [0.94 - 1.05]	0.98 [0.91 - 1.05]	0.25
	Longest slope	3.14e-03 [-4.8e-3 - 0.012]	-3.8e-03 [-9.9e-03 - 7.2e-03]	0.16
	Longest duration	64.50 [56.00 - 76.00]	64.00 [52.00 - 81.25]	0.88
	Steepest slope	0.14 [0.09 - 0.21]	0.15 [0.10 - 0.21]	0.86
	Complete slope	-4.81e-06 [-5.3e-04 - 7.4e04]	-9.37e-05 [-5.2e-04 - 5.2e-04]	0.68

The features are expressed as median and the interquartile ranges [IQR], and are compared using the Mann-Whitney U test. * indicates a significant p-value.

Table A.2: Trend features of the diastolic peak amplitude for deteriorated and non-deteriorated group.

		Deteriorated (n=24)	**Non-deteriorated (n=165)**	**p-value**
Feature	**Trend**			
Diastolic Peak Amplitude	Minimum	-	-	-
	Maximum	-	-	-
	Delta	-	-	-
	Mean	-	-	-
	Median	-	-	-
	Standard deviation	0.40 [0.39 - 0.42]	0.40 [0.39 - 0.41]	0.09
	Coefficient of Variation	0.94 [0.86 - 1.05]	0.87 [0.81 - 0.97]	0.009*
	Number of trend changes	0.06 [0.05 - 0.07]	0.05 [0.04 - 0.05]	0.001*
	Ratio time	1.01 [0.87 - 1.20]	0.94 [0.81 - 1.08]	0.19
	Ratio amount	0.99 [0.96 - 1.05]	0.99 [0.94 - 1.05]	0.51
	Longest slope	-0.01 [-0.02 - 001]	0,00 [-0.01 0.01]	0.38
	Longest duration	60.50 [34.00 - 74.50]	64.00 [52.00 - 81.00]	0.07
	Steepest slope	0.18 [0.09 - 0.27]	0.15 [0.11 - 0.21]	0.82
	Complete slope	-2.07e-04 [-8.3e-04 - 7.0e-04]	1.76e-04 [-3.1e-04 - 7.1e-04]	0.27

The features are expressed as median and the interquartile ranges [IQR], and are compared using the Mann-Whitney U test. * indicates a significant p-value.

Table A.3: Trend features of the pulse width for deteriorated and non-deteriorated group.

		Deteriorated (n=24)	**Non-deteriorated (n=165)**	**p-value**
Feature	**Trend**			
Pulse Width	Minimum	0.20 [0.18 -0.22]	0.21 [0.18 - 0.24]	0.07
	Maximum	0.37 [0.27 - 0.45]	0.41 [0.32 - 0.47]	0.13
	Delta	0.15 [0.10 - 0.23]	0.18 [0.12 -0.23]	0.34
	Mean	0.26 [0.20- 0.32]	0.30 [0.24 - 0.35]	0.11
	Median	0.26 [0.19 - 0.30]	0.28 [0.22 - 0.36]	0.08
	Standard deviation	0.04 [0.03 - 0.06]	0.04 [0.03 - 0.06]	0.42
	Coefficient of Variation	0.13 [0.11 - 0.19]	0.14 [0.10 - 0.19]	0.83
	Number of trend changes	0.02 [0.01 - 0.03]	0.02 [0.02 - 0.03]	0.95
	Ratio time	1.07 [0.71 - 1.51]	1.13 [0.89 - 1.34]	0.57
	Ratio amount	0.99 [0.65 - 1.40]	1.03 [0.89 - 1.27]	0.26
	Longest slope	1.18e-04 [-3.1e-04 - 5.7e-04]	3.1e-04 [-2.9e-04 - 7.6e-04]	0.23
	Longest duration	130.00 [80.50 - 165.50]	107.50 [84.00 - 137.00]	0.54
	Steepest slope	0.002 [0.001 0.01]	0.005 [0.002 0.008]	0.47
	Complete slope	2.41e-05 [-1.4e-04 - 9.9e-05]	3.96e-05 [1.2e-04 - 0.03]	0.03*

The features are expressed as median and the interquartile ranges [IQR], and are compared using the Mann-Whitney U test. * indicates a significant p-value.

Table A.4: Trend features of the Peak to Peak interval for deteriorated and non-deteriorated group.

		Deteriorated (n=24)	**Non-deteriorated (n=165)**	**p-value**
Feature	**Trend**			
Peak to Peak interval	Minimum	0.54 [0.47 - 0.58]	0.60 [0.53 - 0.66]	0.01
	Maximum	0.75 [0.67 - 0.86]	0.80 [0.71 - 0.94]	0.13
	Delta	0.23 [0.18 - 0.27]	0.21 [0.15 - 0.27]	0.43
	Mean	0.64 [0.55 - 0.71]	0.71 [0.62 - 0.81]	0.01*
	Median	0.62 [0.54 - 0.72]	0.70 [0.62 - 0.81]	0.01*
	Standard deviation	0.06 [0.04 - 0.06]	0.05 [0.04 - 0.07]	0.99
	Coefficient of Variation	0.09 [0.06 - 0.11]	0.08 [0.05 - 0.10]	0.24
	Number of trend changes	0.03 [0.02 - 0.05]	0.02 [0.02 - 0.03]	0.26
	Ratio time	1.04 [0.61 - 1.41]	1.23 [0.89 - 1.63]	0.04*
	Ratio amount	1.00 [0.67 - 1.13]	1.11 [0.90 - 1.39]	0.06
	Longest slope	-9.9e-05 [-8.57e-04 - 4.26e-04]	4.98e-04 [-3.31e-04 - 9.02e-04]	0.02*
	Longest duration	90.50 [53.50 - 128.50]	121.00 [84.50 - 165.25]	0.01*
	Steepest slope	5.17e-03 [4.00e-03 - 1.07e-02]	4.44e-03 [2.96e-03 - 8.00e-03]	0.21
	Complete slope	-5.19e-05 [-1.35e04 - 1.37e-04]	8.40e-05 [-3.60e-05 - 1.97e-04]	0.02*

The features are expressed as median and the interquartile ranges [IQR], and are compared using the Mann-Whitney U test. * indicates a significant p-value.

Table A.5: Trend features of the pulse interval for deteriorated and non-deteriorated group.

		Deteriorated (n=124)	**Non-deteriorated (n=165)**	**p-value**
Feature	**Trend**			
Pulse interval	Minimum	0.54 [0.47 - 0.58]	0.60 [0.53 - 0.66]	0.01*
	Maximum	0.74 [0.67 - 0.86]	0.80 [0.71 - 0.94]	0.13
	Delta	0.23 [0.17 - 0.27]	0.20 [0.15 - 0.27]	0.42
	Mean	0.64 [0.55 - 0.71]	0.71 [0.62 - 0.81]	0.01*
	Median	0.62 [0.55 - 0.72]	0.80 [0.62 - 0.82]	0.01*
	Standard deviation	0.05 [0.04 - 0.07]	0.05 [0.04 - 0.07]	0.95
	Coefficient of Variation	0.09 [0.07 - 0.11]	0.08 [0.05 - 0.10]	0.23
	Number of trend changes	0.02 [0.02 - 0.04]	0.02 [0.03 - 0.04]	0.80
	Ratio time	0.98 [0.58 - 1.42]	1.23 [0.92 - 1.61]	0.03*
	Ratio amount	0.93 [0.62 - 1.27]	1.09 [0.89 - 1.42]	0.07
	Longest slope	-4.06e-04 [-9.38e-04 - 5.55e-04]	4.95e-04 [-3.66e-04 - 1.03e-03]	0.01*
	Longest duration	89.00 [55.00 - 105.00]	110.50 [81.50 - 160.50]	0.01*
	Steepest slope	5.33e-03 [3.27e-03 - 1.07e-02]	4.00e-03 [2.67e-3 - 8.68e-3]	0.13
	Complete slope	-5.88e-05 [-1.35e04 - 1.38e-04]	8.73e-05 [-3.63e-05 - 1.98e-04]	0.02*

The features are expressed as median and the interquartile ranges [IQR], and are compared using the Mann-Whitney U test. * indicates a significant p-value.

Table A.6: Trend features of the crest time for deteriorated and non-deteriorated group.

		Deteriorated (n=24)	**Non-deteriorated (n=165)**	**p-value**
Feature	**Trend**			
Crest time	Minimum	0.18 [0.16 - 0.20]	0.18 [0.18 - 0.20]	0.10
	Maximum	0.24 [0.21 - 0.31]	0.24 [0.22 - 0.26]	0.29
	Delta	0.06 [0.03 - 0.12]	0.05 [0.03 - 0.07]	0.16
	Mean	0.21 [0.20 - 0.22]	0.21 [0.20 - 0.22]	0.99
	Median	0.21 [0.20 - 0.22]	0.20 [0.20 - 0.22]	0.78
	Standard deviation	0.01 [0.01 - 0.03]	0.01 [0.01 - 0.02]	0.44
	Coefficient of Variation	0.07 [0.04 - 0.13]	0.06 [0.05 - 0.08]	0.43
	Number of trend changes	0.01 [0.01 - 0.02]	0.01 [0.01 - 0.02]	0.65
	Ratio time	0.79 [0.43 - 1.17]	1.14 [0.69 - 1.76]	0.01*
	Ratio amount	0.87 [0.61 - 1.00]	1.05 [0.83 - 1.27]	<0.01*
	Longest slope	-1.44e-05 [-1.02e-02 - 1.59e-04]	1.48e-05 [-9.25e-05 - 1.09e-04]	0.97
	Longest duration	143.00 [87.50 - 225.75]	169.50 [102.00 - 222.00]	0.41
	Steepest slope	8.00e-04 [6.67e-04 - 3.01e-03]	1.14e-03 [7.82e-04 - 2.43e-03]	0.60
	Complete slope	-5.51e-08 [-4.5e-05 - 5.11e-06]	1.21e-05 [7.06e-06 - 3.37e-05]	0.01*

The features are expressed as median and the interquartile ranges [IQR], and are compared using the Mann-Whitney U test. * indicates a significant p-value.

Table A.7: Trend features of the delta time for deteriorated and non-deteriorated group.

		Deteriorated (n=24)	**Non-deteriorated (n=165)**	**p-value**
Feature	**Trend**			
Delta time	Minimum	0.20 [0.17 - 0.22]	0.22 [0.20 - 0.24]	0.11
	Maximum	0.26 [0.26 - 0.28]	0.28 [0.26 - 0.30]	0.14
	Delta	0.08 [0.06 - 0.10]	0.06 [0.04 - 0.08]	0.05*
	Mean	0.23 [0.22 - 0.25]	0.25 [0.24 - 0.27]	0.01*
	Median	0.22 [0.22 - 0.26]	0.26 [0.24 - 0.26]	<0.01*
	Standard deviation	0.02 [0.01 - 0.02]	0.01 [0.01 - 0.02]	0.08
	Coefficient of Variation	0.08 [0.05 - 0.10]	0.06 [0.04 - 0.08]	0.01*
	Number of trend changes	0.02 [0.01 - 0.03]	0.02 [0.01 - 0.02]	0.12
	Ratio time	0.93 [0.69 - 1.33]	1.06 [0.76 - 1.38]	0.43
	Ratio amount	0.96 [0.83 - 1.20]	1.00 [0.85 - 1.21]	0.82
	Longest slope	-8.55-05 [-2.61e-04 - 2.30e-04]	2.44e-05 [-1.18e-04 - 1.93e-04]	0.23
	Longest duration	99.50 [64.50 - 142.00]	133.00 [92.50 - 185.75]	0.07
	Steepest slope	1.86e-03 [1.26e-03 - 4.00e-04]	1.53e-03 [1.00e-03 - 2.67e-03]	0.15
	Complete slope	-3.48e-06 [-2.77e-05 - 5.07e-05]	5.92e-07 [-1.47e-05 - 2.59e-05]	0.69

The features are expressed as median and the interquartile ranges [IQR], and are compared using the Mann-Whitney U test. * indicates a significant p-value.

Table A.8: Trend features of the infliction point area for deteriorated and non-deteriorated group.

		Deteriorated (n=24)	**Non-deteriorated (n=165)**	**p-value**
Feature	**Trend**			
Infliction point area	Minimum	0.15 [0.04 - 0.27]	0.19 [0.11 - 0.31]	0.07
	Maximum	0.61 [0.32 - 0.73]	0.60 [0.44 - 0.79]	0.52
	Delta	0.38 [0.27 - 0.55]	0.37 [0.26 - 0.53]	0.89
	Mean	0.32 [0.15 - 0.44]	0.37 [0.28 - 0.51]	0.07
	Median	0.30 [0.14 - 0.44]	0.38 [0.27 - 0.49]	0.03*
	Standard deviation	0.09 [0.06 - 0.14]	0.09 [0.06 - 0.12]	0.96
	Coefficient of Variation	0.30 [0.20 - 0.52]	0.24 [0.18 - 0.31]	0.02*
	Number of trend changes	0.06 [0.04 - 0.06]	0.05 [0.04 - 0.05]	0.01*
	Ratio time	0.95 [0.83 - 1.16]	1.13 [0.96 - 1.31]	0.03*
	Ratio amount	0.97 [0.89 - 1.19]	1.03 [0.93 - 1.12]	0.44
	Longest slope	2.05e-04 [-1.74e-03 - 1.75e-03]	1.20e-03 [-1.49e-03 - 2.50e-03]	0.36
	Longest duration	53.00 [46.00 - 73.75]	65.00 [52.75 - 81.25]	0.03*
	Steepest slope	1.99e-02 [1.29e-02 - 3.82e-02]	2.00e-02 [1.27e-02 - 3.62e-02]	0.92
	Complete slope	-2.94e-05 [-2.55e-04 - 2.30e-04]	7.72e-05 [-7.15e-05 - 2.35e-04]	0.12

The features are expressed as median and the interquartile ranges [IQR], and are compared using the Mann-Whitney U test. * indicates a significant p-value.

Table A.9: Trend features of the reflection index for deteriorated and non-deteriorated group.

		Deteriorated (n=24)	**Non-deteriorated (n=165)**	**p-value**
Feature	**Trend**			
Reflection index	Minimum	25.16 [11.00 - 32.54]	26.07 [17.18 - 35.14]	0.21
	Maximum	60.74 [44.13 - 68.10]	60.20 [49.04 - 69.12]	0.66
	Delta	33.90 [28.62 - 40.69]	31.08 [23.80 - 38.19]	0.18
	Mean	40.31 [27.63 - 47.62]	42.73 [33.78 - 51.18]	0.24
	Median	39.50 [29.34 - 47.94]	42.72 [33.43 - 51.30]	0.22
	Standard deviation	8.18 [6.64 - 11.97]	7.10 [5.25 - 9.17]	0.06
	Coefficient of Variation	0.25 [0.16 - 0.37]	0.17 [0.13 - 0.24]	0.01*
	Number of trend changes	0.06 [0.05 - 0.06]	0.05 [0.04 - 0.06]	<0.01*
	Ratio time	0.95 [0.83 - 1.16]	1.13 [0.96 - 1.31]	0.03
	Ratio amount	0.97 [0.89 - 1.19]	1.03 [0.93 - 1.12]	0.44
	Longest slope	-0.08 [-0.24 - 0.13]	0.06 [-0.17 - 0.16]	0.28
	Longest duration	60.50 [45.00 - 75.50]	64.00 [51.75 - 80.50]	0.30
	Steepest slope	2.44 [1.71 - 3.87]	1.82 [1.26 - 2.88]	0.02*
	Complete slope	-3.39e-03 [2.62e-02 - 1.87e-02]	5.73e-03 [-4.87e-03 - 1.69e-02]	0.12

The features are expressed as median and the interquartile ranges [IQR], and are compared using the Mann-Whitney U test. * indicates a significant p-value.

Table A.10: Trend features of the augmentation index for deteriorated and non-deteriorated group.

		Deteriorated (n=24)	**Non-deteriorated (n=165)**	**p-value**
Feature	**Trend**			
Augmentation index	Minimum	39.17 [31.77 - 55.80]	39.58 [30.76 - 50.79]	0.66
	Maximum	74.80 [67.37 - 88.92]	73.89 [64.79 - 82.78]	0.21
	Delta	33.91 [28.63 - 40.59]	30.93 [23.94 - 38.18]	0.18
	Mean	59.69 [52.38 - 72.37]	57.27 [48.82 - 66.22]	0.24
	Median	60.50 [52.06 - 70.66]	57.27 [48.82 - 66.22]	0.22
	Standard deviation	8.18 [6.64 - 11.97]	7.10 [5.25 - 9.17]	0.06
	Coefficient of Variation	0.13 [0.11 - 0.19]	0.13 [0.09 - 0.18]	0.40
	Number of trend changes	0.05 [0.04 - 0.06]	0.05 [0.04 - 0.06]	0.99
	Ratio time	1.00 [0.84 - 1.19]	0.95 [0.81 - 1.11]	0.46
	Ratio amount	1.00 [0.90 - 1.11]	0.97 [0.89 - 1.06]	0.40
	Longest slope	0.13 [-0.16 - 0.24]	-0.09 [-0.15 - 0.18]	0.39
	Longest duration	64.00 [55.00 - 72.00]	61.00 [51.00 - 75.25]	0.94
	Steepest slope	1.73 [1.40 - 3.04]	1.87 [1.34 - 2.86]	0.95
	Complete slope	3.39e-03 [-1.87e-02 - 2.62e-02]	-5.73e-03 [-1.69e-02 - 4.87e-03]	0.12

The features are expressed as median and the interquartile ranges [IQR], and are compared using the Mann-Whitney U test. * indicates a significant p-value.

Table A.11: Trend features of the heart rate for deteriorated and non-deteriorated group.

		Deteriorated (n=24)	**Non-deteriorated (n=165)**	**p-value**
Feature	**Trend**			
Heart rate	Minimum	80.92 [69.49 - 89.72]	74.70 [63.83 - 84.28]	0.13
	Maximum	110.09 [103.45 - 127.66]	100.00 [91.28 - 113.87]	0.01*
	Delta	29.78 [26.15 - 41.67]	25.81 [18.74 - 34.35]	0.01*
	Mean	95.63 [84.94 - 109.50]	85.02 [74.81 - 93.63]	0.01*
	Median	96.77 [83.40 - 109.13]	85.71 [73.17 - 96.77]	0.01*
	Standard deviation	7.64 [6.07 - 10.29]	6.59 [4.45 - 9.05]	0.08
	Coefficient of Variation	0.08 [0.06 - 0.11]	0.08 [0.05 - 0.10]	0.53
	Number of trend changes	0.02 [0.02 - 0.04]	0.03 [0.02 - 0.03]	0.18
	Ratio time	0.96 [0.80 - 1.51]	0.81 [0.63 - 1.07]	0.02*
	Ratio amount	1.14 [0.89 - 1.76]	0.90 [0.73 - 1.08]	0.01*
	Longest slope	0.10 [-0.06 - 0.21]	-0.06 [-0.10 - 0.05]	<0.01*
	Longest duration	90.00 [55.50 - 117.50]	107.00 [77.25 - 159.50]	0.04*
	Steepest slope	0.89 [0.50 - 1.23]	0.74 [0.41 - 1.19]	0.25
	Complete slope	0.01 [-0.02 - 0.02]	-0.01 [-0.03 - 0.01]	0.03*

The features are expressed as median and the interquartile ranges [IQR], and are compared using the Mann-Whitney U test. * indicates a significant p-value.

Table A.12: Trend features of the relative amplitude for deteriorated and non-deteriorated group.

		Deteriorated (n=24)	**Non-deteriorated (n=165)**	**p-value**
Feature	**Trend**			
Relative amplitude	Minimum	1.68 [1.62 - 1.74]	1.69 [1.61 - 1.75]	0.68
	Maximum	2.04 [1.88 - 2.12]	2.01 [1.92 - 2.10]	1.00
	Delta	0.35 [0.21 - 0.41]	0.32 [0.22 - 0.42]	0.98
	Mean	1.82 [1.77 - 1.88]	1.83 [1.76 - 1.89]	0.81
	Median	1.81 [1.76 - 1.87]	1.83 [1.76 - 1.89]	0.81
	Standard deviation	0.08 [0.05 - 0.10]	0.07 [0.05 - 0.10]	0.76
	Coefficient of Variation	0.04 [0.03 -0.06]	0.05 [0.04 - 0.06]	0.08
	Number of trend changes	0.05 [0.05 - 0.06]	0.05 [0.04 - 0.06]	0.08
	Ratio time	1.01 [0.90 -1.20]	0.94 [0.84 - 1.10]	0.17
	Ratio amount	0.94 [0.83 - 1.03]	0.99 [0.93 - 1.05]	0.06
	Longest slope	6.18e-04 [-2.07e-03 - 1.73e-03]	-3.73e-04 [-1.85e-03 - 1.86e-03]	0.60
	Longest duration	62.50 [42.50 - 69.50]	63.00 [48.00 - 74.25]	0.34
	Steepest slope	0.02 [0.01 - 0.03]	0.02 [0.02 - 0.03]	0.06
	Complete slope	-4.03e-05 [-2.51e-04 - 9.33e-05]	2.54e-05 [-6.29e-05 - 1.16e-04]	0.04*

The features are expressed as median and the interquartile ranges [IQR], and are compared using the Mann-Whitney U test. * indicates a significant p-value.

Table A.13: Trend features of the acceleration plethysmography feature for deteriorated and non-deteriorated group.

		Deteriorated (n=24)	**Non-deteriorated (n=165)**	**p-value**
Feature	**Trend**			
APG	Minimum	-1.27 [-1.46 - -1.15]	-1.21 [-1.28 - -1.12]	0.02*
	Maximum	-0.90 [-1.10 - -0.82]	-0.87 [-0.98 - -0.76]	0.14
	Delta	0.38 [0.30 - 0.51]	0.32 [0.25 - 0.41]	0.07
	Mean	-1.08 [-1.21 - -1.00]	-1.04 [-1.13 - -0.95]	0.04*
	Median	-1.07 [-1.21 - -0.99]	-1.04 [-1.14 - -0.94]	0.03*
	Standard deviation	0.09 [0.06 - 0.12]	0.07 [0.06 - 0.10]	0.07
	Coefficient of Variation	-0.09 [-0.11 - -0.06]	0.05 [0.04 - 0.05]	0.01
	Number of trend changes	0.05 [0.04 - 0.06]	0.05 [0.04 - 0.05]	0.01*
	Ratio time	0.97 [0.85 - 1.08]	0.98 [0.82 - 1.17]	0.65
	Ratio amount	0.96 [0.85 - 1.03]	1.01 [0.94 - 1.10]	0.01*
	Longest slope	-4.09e-04 [-3.33e-03 - 2.10e-03]	1.01e-03 [-1.55e-03 - 2.17e-03]	0.20
	Longest duration	63.00 [51.00 - 72.00]	64.00 [52.75 - 77.25]	0.39
	Steepest slope	0.02 [0.01 - 0.03]	0.02 [0.01 - 0.03]	0.69
	Complete slope	1.64e-06 [-3.42e-04 - 7.07e-05]	5.53e-05 [-3.48e-04 - 1.52e-04]	0.03*

The features are expressed as median and the interquartile ranges [IQR], and are compared using the Mann-Whitney U test. * indicates a significant p-value.

Table A.14: Trend features of the stroke volume for deteriorated and non-deteriorated group.

		Deteriorated (n=24)	**Non-deteriorated (n=165)**	**p-value**
Feature	**Trend**			
Stroke volume	Minimum	2.96 [1.55 - 4.30]	3.15 [2.07 - 4.40]	0.49
	Maximum	26.52 [21.11- 34.00]	28.18 [24.71 - 33.03]	0.58
	Delta	23.24 [16.88 - 31.66]	24.92 [21.25 - 29.38]	0.61
	Mean	11.80 [9.89 - 14.10]	13.36 [12.01 - 14.72]	0.01*
	Median	9.81 [9.00- 12.88]	12.81 [11.00- 14.46]	<0.01*
	Standard deviation	5.84 [4.33 - 7.80]	5.74 [4.67 - 6.85]	0.99
	Coefficient of Variation	0.51 [0.40 - 0.69]	0.43 [0.36 - 0.53]	0.06
	Number of trend changes	0.05 [0.04 - 0.06]	0.05 [0.04 - 0.05]	0.16
	Ratio time	1.03 [0.89 - 1.17]	0.91 [0.80 - 1.06]	0.03*
	Ratio amount	0.99 [0.91 - 1.09]	0.99 [0.92 - 1.05]	0.97
	Longest slope	0.02 [-0.06 - 0.16]	-0.05 [-0.13 - 0.12]	0.23
	Longest duration	65.50 [56.50 - 76.50]	65.00 [53.00 - 80.00]	1.00
	Steepest slope	1.89 [1.23 - 3.33]	2.10 [1.50 - 2.94]	0.55
	Complete slope	9.75e-04 [-7.52e-03 - 1.04e-2]	-1.07e04 [-5.46e-03 - 7.40e-03]	0.94

The features are expressed as median and the interquartile ranges [IQR], and are compared using the Mann-Whitney U test. * indicates a significant p-value.

Table A.15: Trend features of the cardiac output for deteriorated and non-deteriorated group.

		Deteriorated (n=24)	**Non-deteriorated (n=165)**	**p-value**
Feature	**Trend**			
Cardiac output	Minimum	263.00 [145.08 - 396.15]	269.55 [178.56 - 402.57]	0.81
	Maximum	2434 [1864 - 3222]	2420 [2081 - 2813]	0.99
	Delta	2135 [1538 - 2984]	2090 [1736 - 2559]	0.92
	Mean	1101 [949 - 1220]	1146 [1047 - 1227]	0.28
	Median	936 [831 - 1144]	1094 [959 - 1192]	0.03*
	Standard deviation	526 [366 - 698]	482 [389 588]	0.50
	Coefficient of Variation	0.50 [0.36 - 0.70]	0.43 [0.35 - 0.53]	0.16
	Number of trend changes	0.05 [0.04 - 0.06]	0.05 [0.04 - 0.05]	0.15
	Ratio time	0.96 [0.80 - 1.11]	0.90 [0.79 - 1.06]	0.21
	Ratio amount	0.98 [0.93 - 1.06]	0.98 [0.91 - 1.04]	0.66
	Longest slope	2.92 [-7.85 - 12.44]	-5.07 [-12.77 - 7.48]	0.17
	Longest duration	64.00 [57.00 - 79.50]	66.00 [54.75 - 80.00]	0.84
	Steepest slope	155.82 [116.65 - 286.01]	176.69 [131.47 - 256.14]	0.70
	Complete slope	0.12 [-0.56 - 0.70]	-0.15 [-0.58 - 0.47]	0.56

The features are expressed as median and the interquartile ranges [IQR], and are compared using the Mann-Whitney U test. * indicates a significant p-value.

Table A.16: P-values of the PPG trend features. The green highlighted p-values are significant (<0.05)

Subfeature	SPA	DPA	PW	PtP	PI	CT	DT	IPA	RI	AI	HR	RA	APG	SV	CO
Min	-	-	0.071	0.008	0.009	0.102	0.107	0.074	0.211	0.665	0.127	0.679	0.021	0.486	0.806
Max	-	-	0.134	0.135	0.130	0.287	0.143	0.519	0.665	0.212	0.009	0.995	0.140	0.580	0.992
Delta	-	-	0.338	0.428	0.421	0.156	0.047	0.887	0.176	0.180	0.013	0.982	0.073	0.605	0.922
Mean	-	-	0.112	0.007	0.007	0.992	0.007	0.070	0.240	0.240	0.007	0.862	0.039	0.015	0.284
Median	-	-	0.083	0.006	0.006	0.778	0.005	0.030	0.219	0.219	0.006	0.806	0.030	0.001	0.033
Standard deviation	0.333	0.098	0.423	0.992	0.947	0.440	0.076	0.957	0.057	0.057	0.078	0.763	0.067	0.986	0.504
Coefficient of variation	0.132	0.009	0.831	0.244	0.235	0.426	0.015	0.016	0.013	0.403	0.532	0.766	0.225	0.062	0.163
Trend changes	0.061	0.001	0.954	0.256	0.803	0.646	0.122	0.011	0.002	0.989	0.178	0.079	0.011	0.158	0.153
Ratio time	0.169	0.188	0.571	0.037	0.026	0.006	0.432	0.027	0.246	0.459	0.019	0.173	0.650	0.032	0.209
Ratio amount	0.246	0.511	0.265	0.056	0.074	0.000	0.819	0.437	0.233	0.396	0.006	0.056	0.013	0.973	0.656
Longest slope	0.164	0.376	0.233	0.015	0.013	0.975	0.233	0.362	0.282	0.392	0.002	0.599	0.196	0.229	0.170
Longest duration	0.883	0.074	0.538	0.011	0.007	0.415	0.072	0.029	0.299	0.943	0.042	0.340	0.391	1.000	0.835
Steepest slope	0.856	0.815	0.473	0.212	0.135	0.595	0.146	0.919	0.024	0.951	0.246	0.059	0.691	0.548	0.697
Slope complete	0.685	0.273	0.033	0.025	0.023	0.007	0.685	0.117	0.116	0.116	0.032	0.040	0.029	0.944	0.561

A.4 P-values PPG features

A.5 Correlation plots

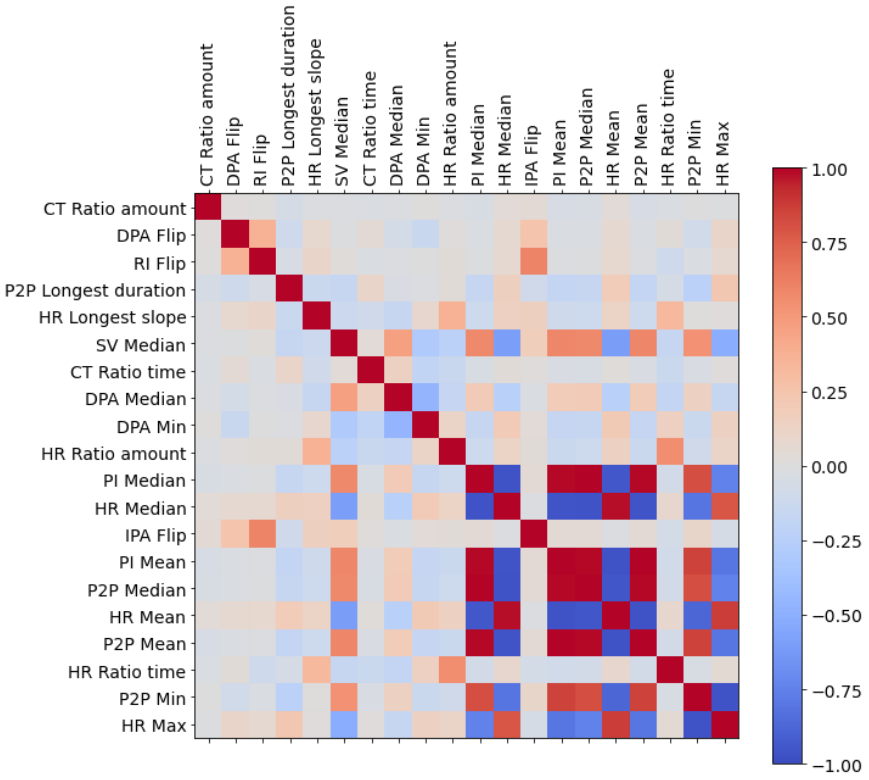


Figure A.3: Correlation plot of best 20 features

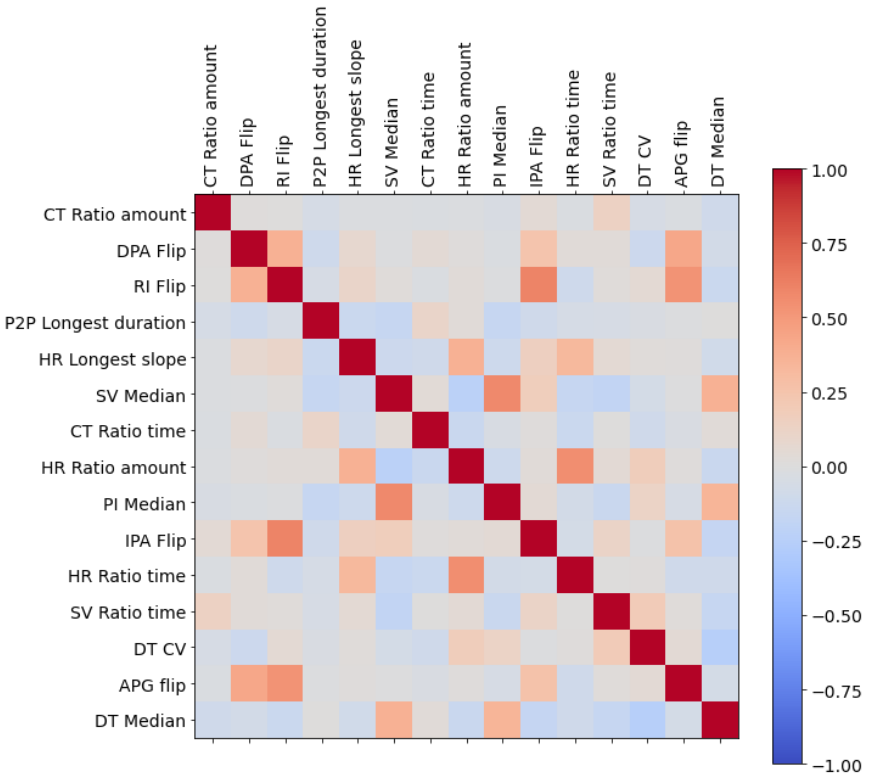


Figure A.4: Correlation plot of best 15 features