ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Prediction of Cerebral Autoregulation in Intensive Care Patients

Master Thesis

A. Kündig

January 26, 2016

Supervisor: Prof. Gábor Székely

Advisors: Dr. Valeria De Luca, Dr. Martin Jaggi

Department of Computer Science, ETH Zürich

Abstract

Traumatic brain injury (TBI) and subarachnoid hemorrhage (SAH) are leading causes of death. Their treatment however usually relies on simple methods which are neither patient- nor disease-specific. Furthermore, current treatment strategies are reactive and based on the observation of the current state of the patient and its clinical context.

To improve the outcome of TBI and SAH patients it was shown that it is important to monitor cerebral autoregulation (CA). Through CA the brain is able to regulate the cerebral blood flow and prevent permanent brain damage. Even though CA itself is not measurable directly, it can be quantified by so called CA indices.

In this work we propose multiple predictive models to forecast the physiological parameters ICP, ABP, and CPP and the CA indices PRx, TF, and IAAC up to two hours into the future. For our proposed models we selected the best out of 9 different sets of feature classes for each prediction horizon and for each prediction target. The different feature classes were derived from statistical, spectral, morphological, and bag of words features.

We evaluated our models on 26 patients from the MIMIC II data set and one 5 patients from a private data set using a leave-one-patient-out cross-validation. For a forecasting horizon of 30 minutes on the MIMIC II data set we achieved a prediction accuracy of 6.67 ± 1.98 mmHg for ABP, 6.87 ± 1.72 mmHg for CPP, 1.94 ± 0.94 mmHg for ICP, 0.28 ± 0.05 for PRx, 0.04 ± 0.04 for TF, and 0.19 ± 0.03 for IAAC. The best models often used statistical summaries, CA indices, or entropy based features. We achieved a relative decrease of prediction error compared to the baseline by up to 11% (13%, 13%) for ICP (ABP, CPP) and 24% (21%), for PRx and IAAC respectively.

Acknowledgements

Contents

Chapter 1

Introduction

According to the Swiss Neurological Society, Traumatic Brain Injury (TBI) is the leading cause of death for people below the age of 44[1]. TBI is in most cases the result of a sudden impact or collision of the head. Typical causes are vehicle accidents, sports injuries, and falls. The initial injury to the head is usually called the primary injury.

However, secondary injuries might occur hours or days after the primary injury and are usually more dangerous than the primary injury. Secondary injuries include damage to the blood-brain barrier allowing bacteria to enter the brain, cerebral edema (cerebral = related to the brain, edema = accumulation of fluid), and cerebral haematoma (haematoma = clotted blood within tissue). More specifically, the cerebral edema and haematoma cause a regional swelling of the brain. This then causes an increase of the intracranial pressure (ICP) (intracranial = within the skull) and hence it increases the resistance for the blood flowing through the brain. The increased resistance then often leads to regional ischemia (under-supply of blood) or hypoxia (under-supply of oxygen) and thus to possible brain damage.

Traumatic brain injury is not the only injury causing an increase in intracranial pressure, ischemia, and hypoxia. A second injury called subarachnoid hemorrhage (SAH) (subarachnoid = below the brain membrane named 'arachnoid', hemorrhage = bleeding) can either occur spontaneously or as a result of a trauma [44], stroke [2], surgery [54], or disorders affecting the blood vessels. The bleeding resulting from SAH causes secondary injuries similar to TBI.

Insufficient supply of oxygen and nutrition to the brain in TBI and SAH patients often results in permanent brain damage. Thus, TBI and SAH are usually associated with bad outcome.

---

[1]http://www.swissneuro.ch/schaedelhirntrauma

A third cause for increased intracranial pressure is a medical condition called Hydrocephalus [31, 20] (hydro = water, cephalus = head). People suffering from Hydrocephalus have an abnormal accumulation of Cerebrospinal Fluid (CSF) (CSF = the fluid below the membrane 'arachnoid' and in the spine) inside the skull. This accumulation leads to a global increase in intracranial pressure and can also cause ischemia and hypoxia.

Patient monitoring and treatment In Neurological intensive care units (NICU) one nowadays still relies on simple guidelines to treat TBI and SAH. The current clinical guidelines issued by the trauma foundation [7] require the doctors to monitor the intracranial pressure and keep it below the threshold of 20 $mmHg$. These guidelines are simplistic, not patient- nor context-specific and might not result in a better outcome. Shafi et al. [51] analyzed the National Trauma Data Bank for the period 1994–2001, analyzing the outcome of patients who had their ICP monitored, and found an increase of mortality of 45% compared to patients who were not monitored. To improve the current guidelines, Lazardis et al. [34] proposed a patient specific threshold for intracranial pressure guided treatment.

Proactive Treatment Current treatment strategies are reactive and based on the observation of the current state of the patient and its clinical context. There is no widely accepted method for predicting the future state. In addition, there is not even wide acceptance for CA monitoring in intensive care. Developing a predictive model and demonstrating its applicability could establish CA monitoring as a necessary indicator in treatment of cerebral injuries and it could change the treatment method from reactive to proactive. Where doctors currently must rely on a patient's history and current state, they could then anticipate future events which were predicted by the model and potentially improve a patient's outcome.

In the following section we will introduce the concept of cerebral autoregulation. We also provide medical background information relevant to our work.

1.1 Medical and Physiological Background

In this section we will provide an overview on the physiological information related to cerebral autoregulation (CA).

The brain has a relatively stable nutrition demand. Since the brain itself has only very little possibility to store energy, this nutrition demand needs to be covered by blood flowing constantly through the brain. The volume of blood flowing through the brain is measured in liters per second and called Cerebral Blood Flow (CBF) [55].

Since the nutrition demand of the brain is relatively stable, the CBF needs to be relatively constant as well. However, the blood flowing into the brain has a dynamic pressure depending on heart beats, respiration, CO_2 concentration in the blood, movement events, changes in blood thickness resulting from medication, e.g. saline injection, and other external causes. The mechanism that compensates for these dynamic changes in pressure and keeps the blood flow relatively stable is cerebral autoregulation.

CA is mainly held by the brain's vasculature [2]. The blood vessels contract or expand to increase or decrease cerebral resistance by means of myogenic, neurogenic, or metabolic mechanisms [45, 10, 41]. By for example increasing the resistance of the blood vessels through contraction, CA can compensate for an increase in blood pressure and keep the CBF approximately constant.

Since, in addition to BP, there are also physiological and medical conditions that influence the ICP, one more generally states that the CA has to compensate for changes in cerebral perfusion pressure (CPP). CPP is simply calculated as the pressure difference between arterial blood pressure (ABP) and ICP.

CA is active within a certain range of CPP, the lower bound called "Lower Level of Autoregulation" (LLA) and the upper bound called "Upper Level of Autoregulation" (ULA). Both are dynamic and depend on the state of the human [45]. An example for a condition that shifts the levels up is chronic hypertension.

Between the LLA and ULA of CPP the CBF increases from 80% to 120% of CBF in the center of LLA and ULA. Below or above those limits flow becomes pressure passive [4, 22], i.e. changes in ABP are transferred directly to changes in ICP. This is mainly because after vessels have contracted to their minimum diameter, they expand again due to the increasing pressure and because after vessels have been in their maximally relaxed state they cannot expand further.

Classic studies have focused on static CA (sCA) which is the long term response (10 to 30 minutes) of CBF to long term changes in ABP [32]. In these studies, measurements of the global CBF are used as surrogate for CA. However, episodes of dangerous hypo- or hyper-perfusion might be overlooked, as these can be observed only over a short period of time.

During the studies of sCA it was measured that the diameters of the main cerebral arteries remain approximately constant under most conditions [23, 40]. Exceptions are combined hypoxia and hypercapnia [46] or inhalation of isoflurane [50], where vasoconstriction and vasodilatation were observed.

---

[2]A vasculature is the blood vessels or arrangement of blood vessels in an organ or part of the body.

If the vessel diameter stays approximately constant one can assume that the flow is proportionate to the flow velocity. Thus, by monitoring for hypoxia and hypercapnia and not using isoflurane for anesthesia it is then possible to measure CBF through cerebral blood flow velocity (CBFV). The CBFV can be measured in one or both of the cerebral arteries using a method called Transcranial Doppler Sonography. Using the measured CBFV it is then possible to infer the current state of CBF, nutrition, and oxygenation[3].

Newer studies on CA have shifted their focus from sCA to dynamic CA (dCA). dCA studies analyze short term changes, sometimes even on a pulse-by-pulse level, in ICP and CBF. Those dynamic changes could originate from the ABP pulse itself or they could originate from other oscillations like breathing which produce oscillations in the CO_2 concentration of the blood.

To make assumptions on the current state of CA the doctors rely on a set of indicator variables. Those variables include static ICP measurement, CBFV measurement, assessment of the patients coma state, and a selection of CA indices. Those are computed from physiological signals like ABP, ICP, CPP, and CBFV and quantify the capability of the brain to autoregulate CBF and the level of oxygenation and nutrition independent from the systemic status. We discuss some core CA indices in the next chapter.

CA can be compromised in case of brain injury. In TBI and SAH patients local swelling due to an edema or haematoma increases local pressure and inhibits the basic vascular mechanisms of CA. It is therefor important to monitor CA to prevent permanent damage to the brain.

1.2 Motivation

In this work we aim to predict basic physiological parameters, i.e. ABP, CPP, and ICP, which are commonly monitored in neurointensive care, and the main CA indices PRx, IAAC, and TF, based on a combination of high resolution physiological signals. We propose a sophisticated machine learning model to forecast these parameters up to two hours into the future. Forecasting these parameters will suggest clinicians an overview of the future patient status, provide early warnings, and hence enable proactive treatment approaches.

---

[3]In our study we do not make use of CBFV measurements since this data is not commonly available. One reason for this is that the Transcranial Doppler device needs to be readjusted from time to time and is sensitive to patient movement.

Chapter 2

Related Work

In this chapter we will review the most common medical indices quantifying cerebral autoregulation (CA). We will start by looking at the oldest methods evaluating the static CA which only quantifies the state of CA over a longer period of time. Then we will progress to more recently proposed indices evaluating dynamic CA, which quantify the state of CA based on short term fluctuations.

In the second part of this chapter we will focus on the task of prediction and discuss different work that has already been done for forecasting various CA related parameters.

Finally, we will list the contributions presented in this work.

2.1 Autoregulation Indices

CA indices originate from the assumption that the cerebral blood flow (CBF) should remain approximately constant even under external influence like increased or decreased blood pressure (hyper-/hypotension) or mechanical activities and movement.

2.2 Static Autoregulation

Initial studies of CA have focused on long term effects on CA, thus named static CA (sCA). They describe the relation between cerebral perfusion pressure (CPP) and cerebral blood flow (CBF).

First work by Lassen [32] suggested that CBF stays constant over a wide range of CPP values. More recent studies have shown that the CBF follows an S-shaped curve which ranges from 80% to 120% of the baseline CBF during normotension [8]. Figure 2.1 shows the characteristical S-shaped curve
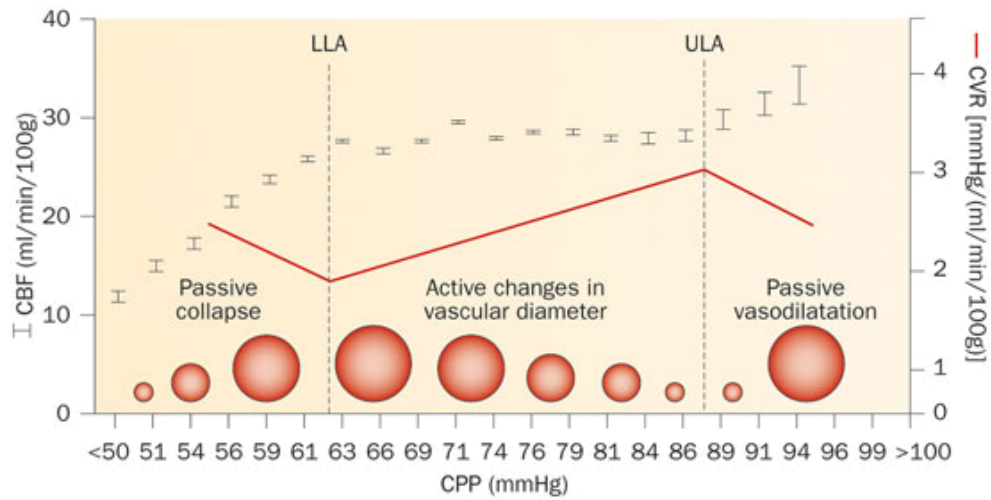
Figure 2.1: S-shaped curve of cerebral autoregulation. Cerebral blood flow stays approximately constant over a wide range of cerebral perfusion pressure between the lower level and the upper level of autoregulation. The corresponding vascular diameters are the reaction of cerebral vasculature to maintain constant flow. (*Figure from Clinical relevance of cerebral autoregulation following subarachnoid haemorrhage by Budohoski et al. [8]*)

of the CPP-CBF interaction, where CBF stays approximately constant between the lower level (LLA) and the upper level (ULA) of autoregulation. Outside of those limits CBF becomes passive to changes in CPP. The figure also shows the vascular diameters corresponding to each CPP value to indicate how the vasculature is able to keep a constant flow in spite of the increased pressure. It is also shown that below the LLA the vessels collapse due to insufficient pressure and above the ULA the vessels dilate because of extreme pressure.

Based on findings of the initial sCA studies, more complex models have been proposed. Gao et al. [22] proposed a compartmental model dividing the cerebrovascular system into compartments with different vessel diameters and then fitted the observed data of sCA studies. The new model had a high accuracy in predicting a patients sCA curve. A major limitations of all sCA models is though, that they require the physician to evaluate the LLA and the ULA before he is able to determine the current position of the patient in the S-shaped sCA curve. Therefore, a static definition of CA is limited in the ICU environment.

In fact, sCA assessments require chemical or mechanical interventions[1] in order to measure CBF for a wide range of CPP values. This is not recom-

---

[1]For example, chemical change of CPP can be induced by medication increasing or decreasing MAP, mechanical change can be induced by tight cuffs or sit to stand maneuvers.

mended for ICU patients who are in critical condition and for whom it is vital to maintain CPP within the limits of CA. Furthermore, the position of a patient in the sCA curve only provides information regarding long term efficiency of CA. It does not quantify how fast CBF returns to a healthy value nor how severe the CA is damaged. Thus, the use of sCA assessment in intensive care units is limited and doctors usually rely on assessment of dynamic CA.

2.3 Dynamic Autoregulation

More recent studies focus on the analysis of dynamic CA (dCA). The change from analysis of static to dynamic CA is mainly possible due to technological advances which have increased the resolution with which we can observe physiological parameters.

One of those technologies is Transcranial Doppler Sonography (TD) [50] which we have already mentioned in the introduction. Using ultrasound and the Doppler Effect, this method is able measure the cerebral blood flow velocity (CBFV) in one or both main cerebral arteries.

An other method for accurately evaluating CBF is Positron Emission Tomography (PET) [9]. However, this method is rarely an option in an intensive care unit since the patient has to be moved to the PET scanner.

dCA studies focus on the reactions of CBF to physiological fluctuation in blood pressure (BP). These spontaneous fluctuations arise from movements, coughing, sleep cycles, heavy breathing, etc. The most prominent BP oscillations during daytime and night time arise in three distinct frequency bands. First, breathing induces oscillations in the respiratory frequency band between 0.2 and 0.4 Hz. Second, variations in vasomotor tone, i.e. contractions of the blood vessels, are present in the band around 0.1 Hz (Meyer waves). Third, very slow and unexplained oscillations are present in the band between 0.02 and 0.07 Hz [42, 43].

Since the reaction of CA to changes in BP is not instantaneous but takes about 5 to 15 seconds, most studies restrict the analysis of CA to slow waves oscillating with less than 0.2 Hz (happening less frequent than every 10 seconds). They assume that oscillatory changes in this frequency band of CBF should be counteracted by a working CA. Therefore, the state of CA is then quantified by the independence of CBF from BP.

Next, we will list relevant dCA indices. We will first start with indices analyzing correlation between different physiological signals of the patient, then we will continue with indices analyzing different spectral properties of the signals, and last we will list some indices analyzing signal morphology.

2.4 Correlation Based Indices

Correlation based indices try to quantify how well CA is working by measuring the correlation between arterial blood pressure (ABP) and an CA related physiological signal, such as intracranial pressure (ICP), cerebral blood flow (CBF), or cerebral tissue oxygenation. They assume that both an increase and a decrease in ABP should result in a reaction of CA and thus often refer to quantifying Cerebrovascular Pressure-Reactivity (CPR) instead of quantifying CA directly.

Based on the available signals and the condition of the patient, different indices have been proposed. Some evaluate cerebrovascular reactivity based on ICP, some based on CBFV, some based on tissue oxygenation. There is a rich set of literature available that compares the applicability of different indices to different medical conditions. However, we restrict our comparison to a core set of CA indices that we later use in our work.

2.4.1 Pressure Reactivity Index

The Pressure Reactivity Index (PRx) [13, 53] proposed by Czosnyka et al. is based on the following intuition: Given the assumption that an increase in ABP should trigger a reaction of CA and we should therefore see a slower increase in ICP, we can correlate the 5-15 second averages of ABP and ICP to see if CA is working. If the correlation coefficient is close to zero or negative, CA is successfully counteracting increases in ABP, if the correlation coefficient is positive, the CA must be degenerate.

To compute the PRx they fist compute the 6 second (sometimes also 10 second) mean ABP and ICP values and then calculate the Pearson correlation of the mean values over the last 3 minutes. Averaging the signals acts as a low pass filter. Hence it is possible to observe CA changes which are longer than 6 to 10 seconds (0.167 to 0.1 Hz and higher). A PRx of less than 0.2 indicates a working CA, while a PRx bigger than 0.4 indicates a degenerated CA. PRx can also be interpreted as phase shift between ICP and ABP waves, where +1 indicates a 0 degree shift and -1 indicates a 180 degree phase shift.

2.4.2 Flow Index

Similar to PRx, the Flow Index (Mx) [11] is computed as the Pearson correlation between 6 to 10 second mean values of CPP and CBFV over the last 3 minutes. While PRx evaluates how well CA can mitigate an increase in ABP, Mx evaluates how strongly an increase in pressure difference influences flow velocity. Thus, it more closely quantifies the effects of ABP on CBF. However, CBFV is relatively hard to measure and is usually not monitored in intensive care units.

2.4.3 Pressure Amplitude Index

A second index related to PRx is the Pressure Amplitude Index (PAx) [47]. PAx is computed as the person correlation between 6 to 10 second mean values of the amplitude of the fundamental first harmonic in ICP derived from the ICP spectrum (AMP) and ABP over the last 3 minutes. PAx directly compares the spectrum derived amplitude with the mean pressure instead of comparing two mean pressures. Therefore, it analyses how strongly ICP pulse amplitude is affected by ABP.

2.4.4 Index of Compensatory Reserve

The Index of Compensatory Reserve (RAP) [3, 31] is closely related to PAx. It correlates AMP with ICP. This index was already published in 1979 and is thus many years older than the other indices listed here.

2.4.5 Single Wave ICP-ABP Amplitude Correlation

The Single Wave ICP-ABP Amplitude Correlation index (IAAC) [17, 18, 19] is a combination and extension of PAx and RAP. It correlates the amplitude of every singe ICP wave over the last 3 minutes with the amplitude of its corresponding ABP wave. Compared to the other indices, this index has shown higher accuracy when correlated with the outcome of patients suffering from subarachnoid hemorrhage. However, this method is computationally more complex because it requires segmentation of the individual ICP pulses. It is less robust to noise in the signal and needs to rely on the correctness of the underlying pulse segmentation algorithm.

2.5 Spectrum Based Indices

The following indices now quantify the state of CA or CPR in the frequency domain or similar representation of the signal.

The main limitation of using the Fourier transformation is, that it assumes the transformed signal to be stationary. Yet, this hypothesis does not always hold under several clinical conditions [39]. Medication, surgery, movement, coughing, and many other factors can have non-stationary effects on the signal. To still be able to compute the indices one thus often assumes that the signals are locally stationary.

2.5.1 Power of Slow Waves

The Power of Slow Waves index [36] index is directly derived from the spectrum of a short time segment of ICP (or CBFV). It is based on the observation that with decreasing levels of CA, the amplitude of waves in frequency

bands below 0.3 Hz increase. This could arise from the fact that ICP itself does not oscillate at low frequencies but only due to external influences. Since CA should be able to counteract oscillations in those low frequencies it is an indicator of a degenerate CA if amplitudes in those frequency bands increase.

2.5.2 Transfer Function Analysis

Transfer Function Analysis (TFA) [57, 58, 16, 48] is a refinement of the analysis of slow waves. Instead of only looking at the magnitude of the oscillations of ICP (or CBFV), TFA estimates how strongly oscillations are transferred from ABP to ICP (or CBFV). The estimation method is based on signal analysis and assumes that CA acts as a high-pass filter on ABP [15, 21]. Thus, TFA assumes that oscillations above a threshold frequency around 0.2 Hz will directly be transferred to ICP (or CBFV) and that oscillations below the threshold should be attenuated. How strongly they are attenuated then indicates how well CA is working.

2.5.3 Wavelet Analysis

Wavelet Analysis (WA) [33] has the advantage of using the Wavelet transformation over using the Fourier Transformation. Thus, one does not need to assume that the input signals are stationary. The result of a wavelet transformation contains information on both frequency and location of a pulse [2]. By computing the Wavelet transformation of ABP and a CA related physiological signal one can compute three measurements of interaction between the two signals: variability of the signals, synchronization which is similar to coherence in TFA, and 'gain' which characterizes amplification of the output signal in comparison to the input signal. Increased gain (with high coherence) may be interpreted as worsening of CA.

2.6 Autoregulation Based Treatment

Current traumatic brain injury (TBI) treatment guidelines do not always lead to an improved outcome. Thus, other treatment guidelines were proposed.

A patient specific CPP threshold was first proposed by Steiner et al. [53]. Others have since then contributed further validation and a similar threshold value for ICP [12, 1, 35]. Steiner et al. found that if they plotted a CA

---

[2] Fourier based methods are also able to provide information on location when one uses the short-time Fourier transform (STFT). However, when using the STFT one needs to trade precision in frequency against precision in time since when adding more samples the Fourier transformation increases precision in frequency because it computes more frequency coefficients but decreases precision in time because it 'averages' the frequency coefficients over a longer period of time.

index value like PRx against CPP over a longer period of time (4h+) the result would usually be a U-shaped curve. They argued that the CPP value at the minimum of that curve is the optimal value for CPP (CPP_{OPT}) to guarantee a working CA. They proposed CPP_{OPT} as the patient specific and context sensitive clinical treatment target. Using the data of 114 head-injured patients, Steiner et al. validated their method by correlating the clinical outcome according to the 6-month Glasgow Outcome Score (GOS) with the deviation of the patient's mean CPP value from CPP_{OPT}. Identification of CPP_{OPT} was possible in 60% of the patients. They showed that if a patient has an average CPP below CPP_{OPT} the GCS would positively correlate with the difference ($r = 0.53, p < .001$), and if a patient's CPP was bigger than CPP_{OPT} GCS would negatively correlate with the difference ($r = -0.40, p < .05$).

Figure 2.2 shows the curve of a second-order polynomial fitted to the CPP-PRx interaction. The CPP_{OPT} is clearly visible at the minimum of the curve at a CPP value of 70 mmHg.

The difficulty with this method is that usually the mean CPP can only be observed within a certain small range because the observation period is too short (2–4 hours). The resulting curve is then often flat or concave and no clear minimum can be computed. The observation period could be increased but the clinical relevance would diminish because, especially in the early period after admission to the intensive care unit, the doctors would not have a threshold value available.

2.7 Predictive Models

Prediction of Intracranial Hypertension

Previous work by Hüser et al. [28, 29] proposed a model to predict intracranial hypertension. They trained and evaluated the model using the publicly available MIMIC II data set and the BrainIT data set. Two main contributions of their work were the analysis of signals at different time scales and the construction of complicated features based on those different time scales.

To build the different time scales, the input signals were first preprocessed, then resampled, and finally stored for feature construction. In a second stage, segments of different length were taken from the resampled signals and used to construct many statistical and morphological features. Each feature had a specific time resolutions and segment length in minutes associated with it. Thus, the same statistical property, e.g. the mean or trend, could be computed for many different scales and segment lengths.

The authors validated their model using 25 records from the MIMIC II data set and 3 records from the BrainIT data set. Doing a 10-fold patient-stratified
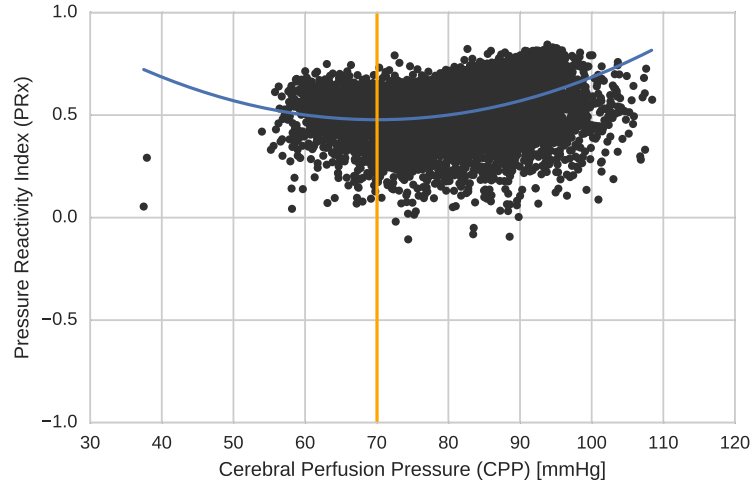
Figure 2.2: The curve of a second-order polynomial fitted to the CPP-PRx interaction in a recording segment of 3 hours. The CPP_{OPT} is clearly visible at the minimum of the curve at a CPP value of 70 mmHg and marked with a vertical line.

cross-validation the authors reported an AUC-ROC score of 0.81 when predicting intracranial hypertension onset events 10 minutes into the future.

Hüser et al. referenced other authors, who have also proposed predictive models for forecasting intracranial hypertension. Guiza et al. [25] reported an AUC-ROC score of 0.87 when predicting intracranial hypertension onset events 30 minutes into the future on a data set of 264 TBI patients. Their model was built based on summary statistics, signal clusterings, frequency-domain analyses and correlations between ICP and ABP of 4 hour minute-by-minute recordings of ICP and ABP. They also included clinical information into their model. Hamilton et al. [26] and Hu et al.[27] build their forecasting model using morphological features. Those features are derived from the segmented ICP pulse shape and contain the location of the three subpeaks, amplitudes, turning points, and latency. Hamilton et al. [26] reports a specificity of 75% coupled with a sensitivity of 90% for a forecasting horizon of 5 minutes on a private data set. Hu et al.[27] reports 99.9% specificity and 37.5% sensitivity on a private data set without TBI patients.

All three authors used physiological signals to predict onsets of intracranial hypertension. They already achieved high scores and also include morphological features into their model. However, they are missing information on cerebral autoregulation which could be an important indicator of a near future onset of intracranial hypertension. CA indices captures the state of CA and thus indicate when the brain is not able anymore to regulate ICP.

Noninvasive Prediction of Mean Intracranial Pressure

Hüser et al. [28, 29] also proposed a predictive model for non-invasively estimating ICP. This predictive model was based on the same multi-scale multi-history feature construction framework and used the same set of features used for forecasting intracranial hypertension. However, they excluded features based on cerebral signals. The authors validated their model using 25 records from the MIMIC II data set and 3 records from the BrainIT data set. Doing a 10-fold patient-stratified cross-validation the authors reported a mean absolute error 3.84 mmHg when non-invasively forecasting ICP.

Similar work was done by Kashif et al. [30]. They proposed a model-based approach requiring no calibration or training on a set of reference patients. Their model used 60-beat segments of ABP and Transcranial Doppler readings of CBFV to estimate the current ICP and is specified in terms of an electrical circuit. They evaluated their model on a set of 37 patients with TBI on which they reported a bias of $1.5mmHg \pm 5.9mmHg$. The advantage of their model is, that it does not need calibration. However, the reported variance of their estimation error seems to indicate that their are not fully capable of capturing all important features.

Invasive Prediction of Mean Intracranial Pressure

Zhang et al. propose an artificial neural network based intracranial pressure mean forecast algorithm [56]. Their proposed online algorithm is based on an artifical neural network (ANN) coupled with an auto-regressive (ARMA) model. They split the continuous time series up into windows of a pre-defined length and then dynamically segment those windows to compute statistical features like mean and standard deviation. The computed features are then given to the ANN-ARMA model for prediction of future ICP means.

Their best model had an R^2 score of 0.93 ± 0.05 (0.81 ± 0.11, 0.56 ± 0.25) for the time horizon $T = 15min$ ($30min$, $45min$). They also report and MSE of $0.88mmHg \pm 0.58mmHg$ ($3.26mmHg \pm 1.96mmHg$, $8.12mmHg \pm 4.72mmHg$) and an RAE of $9\% \pm 3\%$ ($24\% \pm 11\%$, $49\% \pm 23\%$) respectively. They report a very low prediction error but unfortunately they predict the ICP mean value for the full 45 minute window into the future. It would have been more informative to predict the 1 minute mean ICP 45 minutes in the future because extreme events are otherwise averaged out.

2.8 Conclusion

Compared to the presented related work, this work makes the following contributions:

- Prediction of 30-second mean ABP, ICP, CPP, PRx, TF, and IAAC for a forecasting horizon of up to 120 minutes.

- Evaluation of the effect of an increasing time horizon on the selection of features used in the model for prediction.

- Evaluation of the relevance of feature categories with respect to their prediction target and horizon.

Chapter 3

Data Sets

For this work we relied on two data sets, both composed of multiple physiological signals in high resolution and in some cases also clinical information.

3.1 MIMIC II

The public Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) II [1] database [49, 24] contains data collected at the Beth Israel Deaconess Medical Center, a tertiary teaching hospital in Massachusetts. It contains recordings of about 23'000 hospital stays. Those recordings have been anonymized, are publicly available without restrictions, and contain recordings of many physiological parameters including blood pressure (arterial, venous, and other) intracranial pressure, heart rate, breathing frequency, blood oxygenation, and other parameters. The data set is split into wave form data sampled at 125 Hz and numeric data collected or computed every second. Which signals are available is dependent on the decisions made by the ICU staff. Thus, a record only contains recordings of signals which were considered clinically relevant during the time of treatment. As a result, the availability of signals varies heavily.

Out of the 23'000 records only 26 passed our evaluation criterion which required the record to have ICP and ABP signals available in at least 25% of the total recording time. The 26 records resulted in approximately 50 days of recording.

3.1.1 Signals

Individual statistics on the availability of signals and numerics in the selected 26 records are shown in Figure 3.1 and Figure 3.2 respectively. We

---

[1]https://www.physionet.org/mimic2/

fused variations of the same signal like systolic, mean, and diastolic pressure since they are always available together.

Here we list a summary of the different signals and how they are described on the MIMIC II website:

Wave Form

RESP uncalibrated respiration waveform, estimated from thoracic impedance
PLETH uncalibrated raw output of fingertip plethysmograph
ECG (electrocardiographic) waveforms include: AVF, AVL, AVR, I, II, III, MCL, MCL1, V (unspecified precordial lead), V1, and V2
BP (continous blood pressure) waveforms include:
    ABP arterial blood pressure (invasive, from one of the radial arteries)
    ART arterial blood pressure (invasive, from the other radial artery)
    CPP cerebral perfusion pressure
    CVP central venous pressure
    ICP intracranial pressure

Numerics

BP blood pressure (systolic, diastolic, and mean)
HR heart rate
RESP Respiration rate
SpO2 oxygen saturation (from fingertip plethysmography)
TEMP Temperature

In this work we use the signals ABP, ICP, and II and we used the numerics HR, RESP, and SpO2. We selected them because they are available in all 26 records and because they are recorded for almost the total recording time.

3.1.2 Data Access

The MIMIC II data set can conveniently be downloaded as one CSV file per record using the *rdsamp* tool, which is part of the wfdb toolkit [24]. The tool automatically fills the respective columns with NaN if a certain signal is not available for a given time stamp. The waveform signals are sampled at 125 Hz and the numeric signals are sampled at 1 Hz.

3.1.3 Targets

Tables A.1, A.2, and A.3 in the appendix show the percentage of the total recording time for which each target was available. The percentage is listed for each record number for the three selected prediction horizons of 30 minutes, 1 hour, and 2 hours.
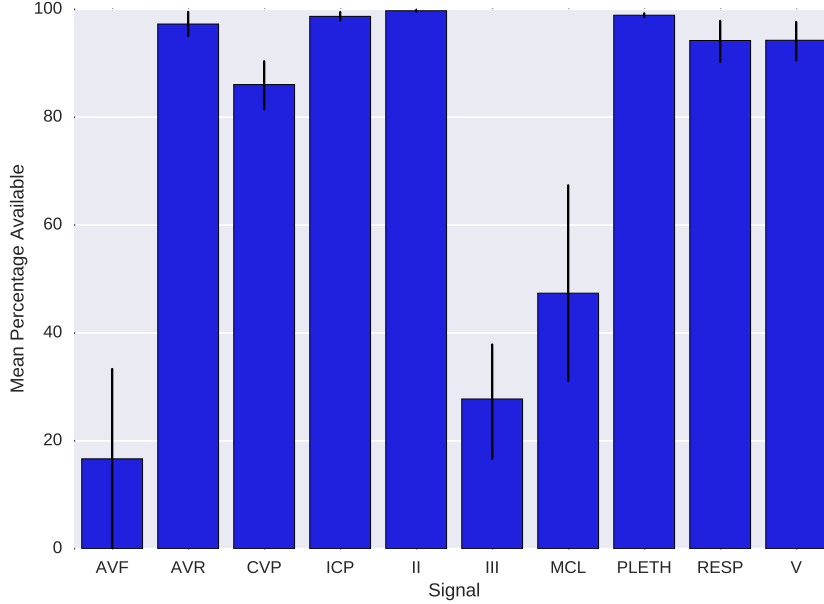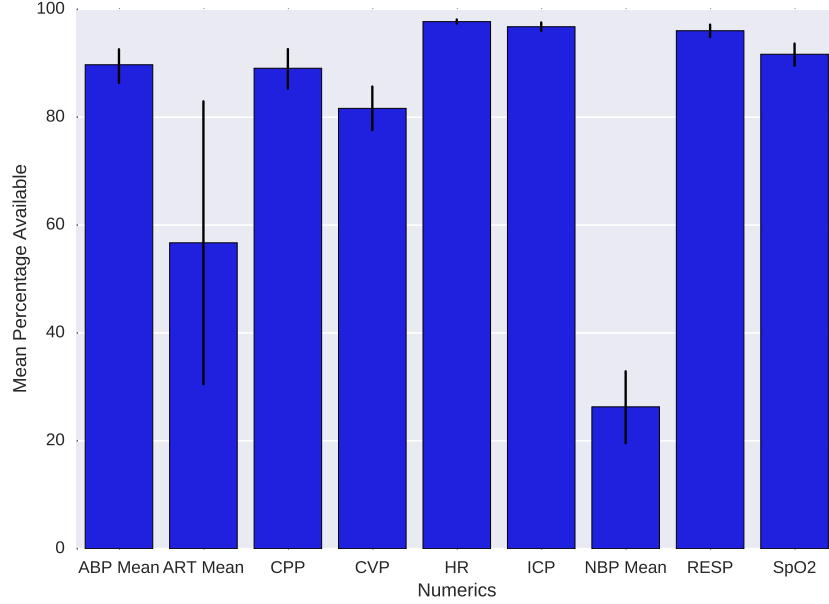
Figure 3.1: Signal availability statistics for the MIMIC II data set. The main signals used in the experiments are: Arterial Blood Pressure (ABP) Intracranial Pressure (ICP) Cerebral Perfusion Pressure (CPP)

The presence of gaps in the considered signals reduced the number of prediction targets, e.g. to compute indices like PRx one needs continuous measurements of input signals over longer periods of time. It is thus possible that after a gap some targets are already available but others are not.

3.2 Cambridge

The Cambridge data set is a private data set collected at the Division of Neurosurgery in the Addenbrooke Teaching Hospital in Cambridge, UK. We had access to a subset of 11 records out of the whole data set. Since some records contained no ECG recording we selected the 5 records which had this data available.

3.2.1 Signals

Figure 3.3 summarizes the availability of the signals for the two sets of records. Compared to the MIMIC II data set, this data set contains no information on SpO2, respiratory frequency, and heart rate. Thus we were constricted in the set of features we could compute from the available data.

17

Figure 3.2: Numerics availability statistics for the MIMIC II data set. The main signals used in the experiments are: Heart Rate (HR) Oxygen saturation in blood (SpO2) Respiratory Frequency (RESP)

3.2.2 Preprocessing

The Cambridge data set records were split into multiple segments stored in CSV format. We concatenated these segments into one CSV file containing the whole record. Similar to the MIMIC II data set, missing values were replaced by NaN. We considered time stamp gaps which occurred within segments and between segments and which were bigger than 5 minutes as missing. Smaller gaps were ignored.

The signals of each record were sampled at a consistent sampling rate in the range $[30, 200]$ Hz. To make the data set comparable with the MIMIC II data set we resampled the records to 125 Hz.

Resampling consisted of the following steps: up-sampling by zero-padding to a sampling rate which is a multiple of 125 Hz (1000 Hz for 200 Hz and 750 Hz for 30 Hz); low-pass filtering to remove aliasing effects; multiplying the signal times the up-sampling factor because zero padding and filtering effectively divided the samples by the up-sampling factor; down-sampling by simply picking every n-th sample where n is the down-sampling factor. We resampled each signal of each segment of each record independently for memory efficiency.
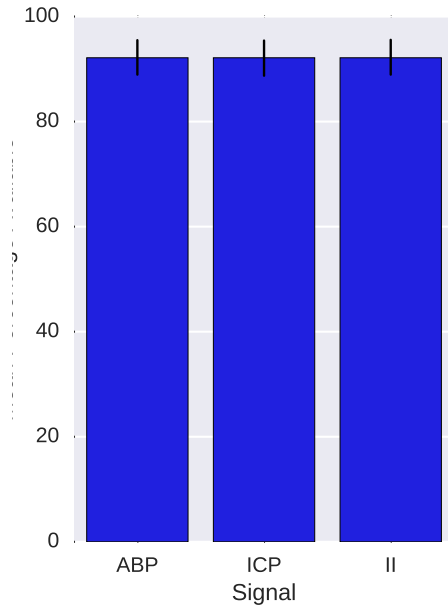
Figure 3.3: Signal availability statistics for the Cambridge data set. CPP can be computed from ABP and ICP.

Since gaps in time stamp values could occur even within segments, we applied the resampling process only to continuous sub-segments within the segments.

3.2.3 Targets

The same information on target availability as for MIMIC II can be taken from Tables A.12, A.13, and A.14. Since the signals sometimes contain gaps, we were not able to compute the target value for the whole time period.

Chapter 4

Methods

We modeled all forecasting problems as regularized linear regression problems. Since we assumed that the relation between the current state and the state at the prediction horizon is non-linear we also included non-linear features into the regression models. The epsilon insensitive absolute loss of the models was then minimized using stochastic gradient descent (SGD).

We built our models from different classes of features each derived from the same initial data. Each feature class focused on a different aspect of the data. The first feature class contained raw input data, the second feature class contained the statistical summaries of the data, the third contained spectral features, the fourth contained Bag of Words features, the fifth contained morphological features, and the sixth contained medical indices. We also used transformed versions of the statistical summary class and the class of medical indices as the seventh and the eighth class. Finally, we include the features proposed by Hüser et al. [28] as the ninth feature class.

Each feature set contained features computed at different time scales. Each time scale was defined as a sampling frequency and a window length. It was therefore possible to compute the same feature over short- and long term windows of time and thus to capture both short- and long-term effects.

We implemented the whole pipeline — preprocesssing, feature construction, target construction, and learning — using an extended version of the Python framework proposed by Hüser et al. [28].

4.1   Preprocessing

Each signal was preprocessed independently, as done in [28]. The main steps of preprocessing were: marking of invalid segments; imputation of missing data via linear interpolation; low-pass filtering to remove high frequency noise and band-pass filtering to remove baseline drift.

Details can be found in [28]. In the proposed work, we modified the signal filtering by switching from a finite impulse response (FIR) based filter (Kaiser) to an infinite impulse response (IIR) based filter (Butterworth). This was mainly done to speed up filtering[1]. In addition, we kept the filter state in memory between the windows which leads a more accurate filtering of consecutively valid windows.

Since the feature construction phase is an online process, preprocessing is also done on batches of input signals.

We changed how signals are filtered by switching

4.2 Features

We used the same set of features proposed by Hüser et al. [28]. In addition, we introduced: a trace based feature, two feature sets based on wavelet transforms, two feature sets based on six auto-regulation indices, and a bag-of-words feature based on a symbolic encoding of signals. The features used in our predictive model are described in the following.

4.2.1 Statistical Summaries

We denote the feature set of statistical summaries as F_{stats}. It contains the input's min, max, mean, median, and slope as well as the variance, standard deviation, skewness, kurtosis, and norm. A similar set of statistical features including some more complex measures of information content like sample entropy were already proposed by Hüser et al. [28].

The minimum and maximum values define the bounds of the input vector $\mathbf{x} = [x_1, \ldots, x_n]$.

$$x_{min} = \min_i x_i \qquad\qquad x_{max} = \max_i x_i$$

The mean and median describe the the location of the input.

$$x_{mean} = \frac{1}{n} \sum^i x_i$$

$$x_{median} = \left\{ \begin{array}{ll} x_{\lceil n/2 \rceil}, & \text{for } n \text{ is odd} \\ \frac{x_{n/2} + x_{1+n/2}}{2}, & \text{for } n \text{ is even} \end{array} \right\}$$

---

[1]Results on the achieved performance gains can be found in Figure A.10 in the appendix.

The slope describes the tendency of the input which can be important to identify drift in a signal.

$$x_{slope} = \beta : \quad \alpha, \beta = \operatorname*{argmin}_{\alpha,\beta} \sum_{i=1}^{n} [x_i - \beta * i - \alpha]^2$$

We define the k-th uncentralized moment as

$$x^{(k)} = \frac{1}{n} \sum^{i} (x_i)^k.$$

From this we also compute several descriptions of the shape of the input distribution and we compute the norm as a measure of energy contained in the input.

$$
\begin{aligned}
x_{var} &= x^{(2)} - [x^{(1)}]^2 \\
x_{std} &= \sqrt{x_{var}} \\
x_{skew} &= \frac{x^{(3)}}{[x^{(2)}]^{3/2}} \\
x_{kurtosis} &= \frac{x^{(4)}}{[x^{(2)}]^2} \\
x_{norm} &= \sqrt{x^{(2)}}
\end{aligned}
$$

4.2.2 Discrete Fourier Transformation

Based on the results obtained in [36] we propose a set of features derived from the Fourier transform of the input. Instead of using the raw Fourier transform as in [28], we compute an estimate of the power spectral density (PSD) using Welch's method. In Welch's method, the spectral density is estimated by moving a sliding window h over the input vector x and computing the periodogram of each window (the windows overlap by m points). All resulting periodograms are then averaged to get the estimate of the power spectral density.

This method has two parameters, where the first parameter h defines the window function and implicitly also the length of the moving window, and the second parameter defines the percentage of overlap of the different windows.

For input signals with frequency $f_{in} = 125Hz$ we chose an overlap of 50% and the hanning window of length $l = 512$. We chose the hanning window because it smooths discontinuities at the boundary of the samples. Using 50% overlap is a compromise between accuracy and estimation of the PSD while not overcounting samples when using the hanning window. Using 512 samples is a compromise between time- and frequency-resolution. These parameters lead to a frequency resolution of $f_{out} = \frac{f_{in}/2}{l/2}Hz = 0.24Hz$, which is sufficient to capture the power of small waves as mentioned in [36] but does not lead to too many coefficients.

To reduce the number of coefficients, we cut off the power spectrum at coefficient $k = 55$. This results in a cutoff frequency of $f_{cut} = k * f_{out} = 13.2Hz$ which is at least 4 times the fundamental frequency of the highest assumed heart rate of 200 beats per minute (3.33 Hz). An example power spectrum can be seen in Figure 4.1. There, the fundamental frequency of the heart rate is clearly visible at around 1.5 Hz (90 beats per minute).

For input signals sampled at $f_{in} = 1Hz$ we also selected an overlap of 50% and used a hanning window of length $l = 16$. Since we decided to keep the resulting PSD vector without cutoff, the resulting frequency resolution is $f_{out} = 0.06Hz$ and a maximum frequency is $f_{cut} = l * f_{out} = 0.5Hz$.

We computed the PDS over segments of 30 seconds since we assumed that most physiological signals are stationary over such a short period of time.

Finally, we also include a statistical summary of the full PSD, see Section 4.2.1.
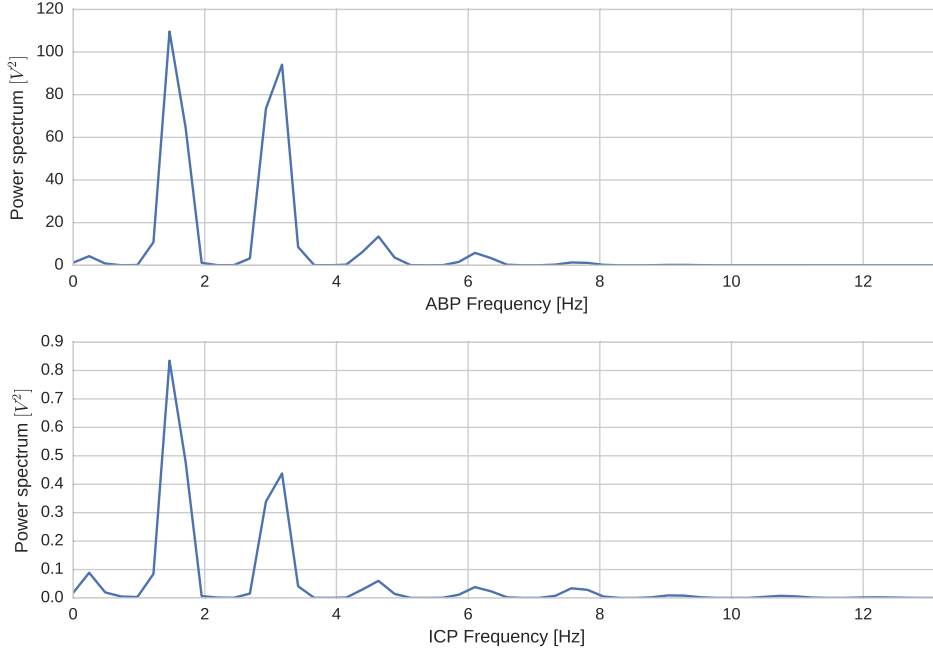
Figure 4.1: The first 55 values of the power spectrum of ABP and ICP sampled at 125 Hz over a time frame of 30 seconds. The fundamental frequency of the heart rate is clearly visible in both spectra at around 1.5 Hz.

4.2.3 Discrete Wavelet Transformation

Based on the results obtained in [33] we propose a set of features derived from the discrete wavelet transform [14] of the input signal. We chose the Daubechies wavelet family for the transformation since it is often used in signal processing tasks. We then computed the complete decomposition of the input signal in the following form

$$wd = [a_k, d_k, d_{k-1}, d_{k-2}, .., d_2, d_1],$$

where a_i and d_i are the coefficients of the approximation and details at level i respectively, with $i \in [1, k]$ and $k = \lceil \log_2 n \rceil$.

At level 1 the input signal of length n is decomposed into a_1 and d_1, each of length 2^{k-1}. Each level a_{i+1} and d_{i+1} is then computed from its predecessor a_i until the length of a_i is 1 at level k. Since the number of samples at any level i is not necessary a power of two but the Daubechies wavelet at level i have 2^{k-i} coefficients we symmetrically expand a_i and x at the boundary if they do not have 2^{k-i} elements.

(a) Arterial blood pressure (ABP).   (b) Intracranial pressure (ICP).
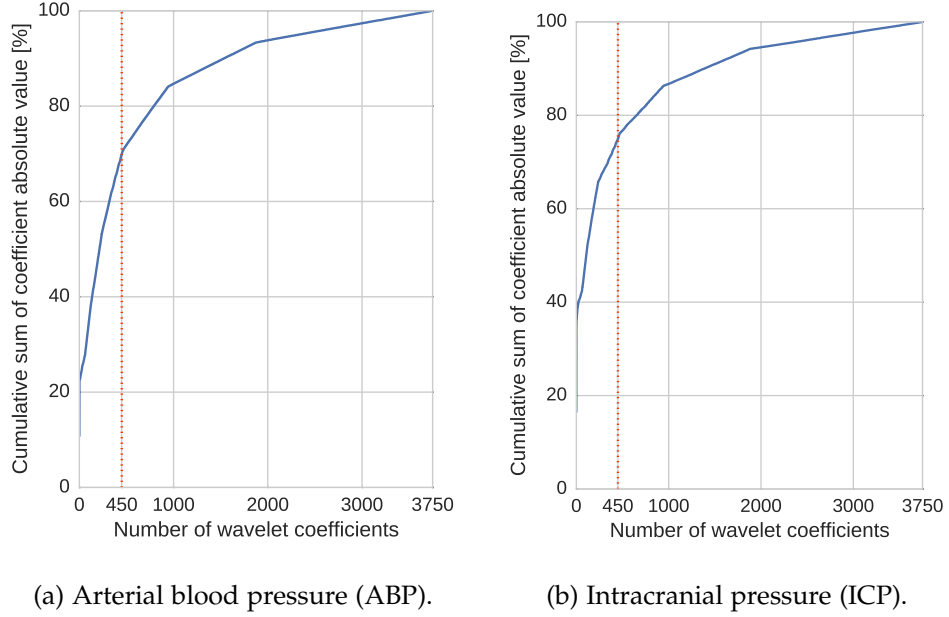
Figure 4.2: The cumulative absolute sum of absolute wavelet coefficients for a sample segment of ABP and ICP containing 3750 sample over a time of 30 seconds each. The cutoff at 450 coefficients is marked with a vertical line.

We thresholded the number of coefficients in *wd* by considering the first $k = 450$ coefficients summing to approximately 70% of the cumulative sum of absolute values. We only thresholded *wd* for signals sampled at 125 Hz since for smaller sample rates the number of coefficients was sufficiently low. In Figure 4.2 an example of the threshold computation for ABP and ICP over a 30 second window is shown.

From the coefficients obtained by the wavelet decomposition we also computed a statistical summary as a set of features, see Section 4.2.1.

4.2.4 Autoregulation Indices

The novelty of the proposed features compared to [28] is the introduction of CA indices. In the specific, we implemented the indices PRx (Section 2.4.1), PAx (Section 2.4.3), RAP (Section 2.4.4), IAAC (Section 2.4.5), SLOW (Section 2.5.1), and TF (Section 2.5.2). We selected the aforementioned indices, as these encode different aspects of the CA state and correlate differently to ICP. An example of these CA indices is shown in Figure 4.3.

PRx We computed x_{PRx} as the Pearson correlation of the last $k = 60$ mean-ABP and mean-ICP values computed by a sliding window of length $w =$

10$seconds$ with step size w, so without overlap.

PAx   We computed x_{PAx} as the Pearson correlation of the last $k = 60$ mean-ABP values and the last k AMP values computed by a sliding window of length $w = 10seconds$ with step size w, so without overlap. AMP is defined as the largest coefficient of the Fourier transformation of the sliding window in the physiological range from $f_{low} = 3Hz$ (20 beats per minute) to $f_{high} = 0.3Hz$ (200 beats per minute). Thus

$$AMP = \max_i |FFT(ICP)_i| \text{ s.t. } f_{low} \leq frequency_i \leq f_{high}$$

RAP   We computed x_{RAP} similar to x_{PAx} but correlating mean-ICP with AMP.

IAAC   We computed x_{IAAC} by first computing the pulse segmentation P of ICP based on the II ECG signal. For each ICP pulse $p \in P$ we then determine the ICP amplitude $A_{icp,p} = \max ICP_p - \min ICP_p$ and the ABP amplitude $A_{abp,p} = \max ABP_p - \min ABP_p$. Finally, we computed the Pearson correlation between $A_{abp} and A_{icp}$.

SLOW   We implemented x_{SLOW} by computing the power spectral density (PSD) of the segment using Welch's method and then summing up the absolute values of the first coefficients up to a cutoff frequency $f_{SLOW} = 0.3Hz$, as mentioned in [36].

TF   For computing x_{TF} we broke down the three components of the transfer function [57, 58, 16, 48] into one number. The first component of the TF is coherence. This is a value in the range $[0, 1]$ that quantifies how well the output signal can be described by a linear function of the input signal. The second component of the TF is phase shift. It defines how much the phase of a particular frequency is shifted in the output signal relative to the same frequency in the input signal. Therefore, the value of phase shift is in the range $[0, 2\pi]$. A high phase shift is generally considered good. The third component of the TF is magnitude. It defines the amplification of a frequency from input signal to output signal. All those three properties are defined for all frequency bands captured by the TF.

We wanted our TF index to be in the range of $[0, 1]$ to make it comparable to other auto-regulation indices proposed in literature. We thus normalize phase shift to the range $[0, 1]$ by dividing it by 2π. Since coherence is in the range $[0, 1]$ but high coherence coupled with low phase shift is considered bad we invert the value of coherence such that low phase shift coupled with low coherence results in a low value. Finally, since magnitude is not bounded and we could not find concrete threshold values to classify the

magnitude as good or bad we omitted magnitude in the computation of the TF index. To conclude, for breaking those three properties of the TF down to one index number we did the following. First, we computed the auto-spectrum and the cross-spectrum for both input signals for the whole segment using Welch's method. Then we normalized the phase shift φ by dividing it by 2π. Next, we multiplied the normalized phase shift with the inverted coherence such that a high inverted coherence and a high normalized phase shift results in high index value. Finally we computed the mean of all the values up to the cutoff frequency of 0.3 Hz. The full computation for the TF index is thus

$$TFIndex_{x,y} = \frac{1}{cutoff} \sum_{i=1}^{cutoff} \frac{\varphi_i}{2\pi} * (1 - coh_i)$$

where:

$$cutoff = \max_i i \quad \text{s.t. } frequency_i < 0.3 \text{ Hz}$$
$$S_{ab} = CrossSpectrum(a, b)$$
$$coh = \frac{abs(S_{xy})}{\sqrt{S_{xx} * S_{yy}}}$$
$$H = \frac{abs(S_{xy})}{S_{xx}}$$
$$\varphi = angle(real(H), imag(H))$$

The frequency resolution and the maximum frequency of a TF is determined by the sampling rate of the input and output signal and the number of samples used for the analysis.

When computing a transfer function one implicitly assumes that the input and output signal are stationary because the analysis makes use of a Fourier transformation. Thus, we decided to compute the transfer function index on segments of length 10 minutes. This is in line with the other indices which are also computed over 10 minute segments.

As an additional feature we also computed the trend of each index over the last 20 minutes. For this, we split the 20 minute segment up into subsegments of 10 minutes with an overlap of 9 minutes and computed the index value on each subsegment for each CA index. We then computed a least squares regression to determine the slope of the resulting index values over the 20 minutes period.
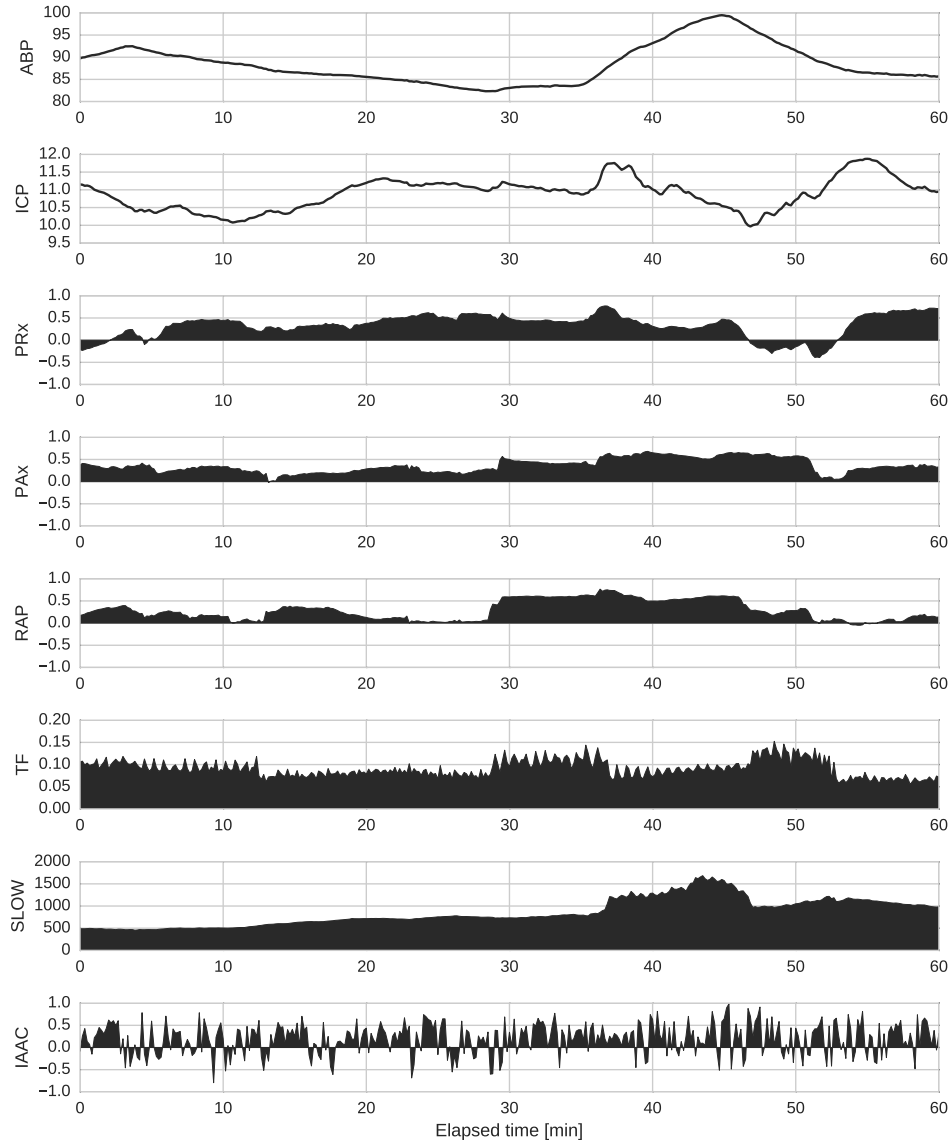
Figure 4.3: Comparison of the different cerebral autoregulation indices computed on a representative 1 h segment. Each index is computed over a sliding window of 10 minutes with 10 seconds step.

4.2.5 SAX Encoded Bag of Words

Based on the work by Lin et al. [38] we proposed a Bag of Words (BoW) based feature. To compute a BoW on a time series segment of continuous values one first needs to discretize the segment. This is sometimes also called 'converting the time series into its symbolic representation'. Like Lin et al. we converted the time series into its symbolic representation using the Symbolic Aggregate approXimation (SAX) encoding [37].

A time series is encoded into its SAX representation based on two parameters, the number of symbols α and the word length w. The time series is first normalized by subtracting its mean and dividing by its standard deviation. Then, the time series is aggregated by computing the piece-wise average of all non-overlapping subsegments of length w. Finally each aggregate is looked up in a table discretizing the whole range of values into α symbols. When computing this lookup table one tries to assign each symbol approximately the same probability. For this one assumes that the time series is normal distributed after normalization. See Figure 4.4 for an example.

To compute the lookup table with α entries one divides the range $(0, 1)$ up into α equally sized segments. Then, for each inner boundary, one computes the inverse cumulative distribution function of the normal distribution. The resulting values represent the upper boundaries of the first $\alpha - 1$ symbols. The last symbol gets assigned to all values larger than the largest boundary.

To compute the BoW from the symbol series, we then need to specify a third parameter, the dictionary word length ω. The resulting dictionary will have α^ω possible words. Since the dictionary size increases exponentially, we need to keep both parameters small. As mentioned by Lin et al., the parameter α does not have such a big effect and can be kept small. The
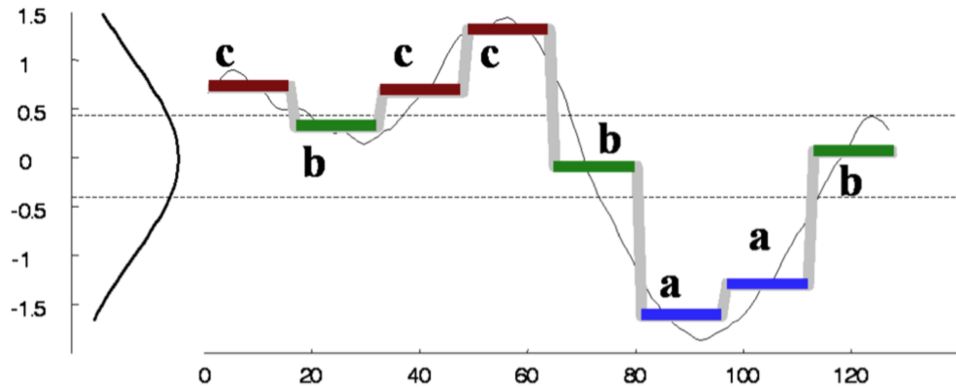
Figure 4.4: An example encoding of a small time series using SAX taken from Lin et al. [38]
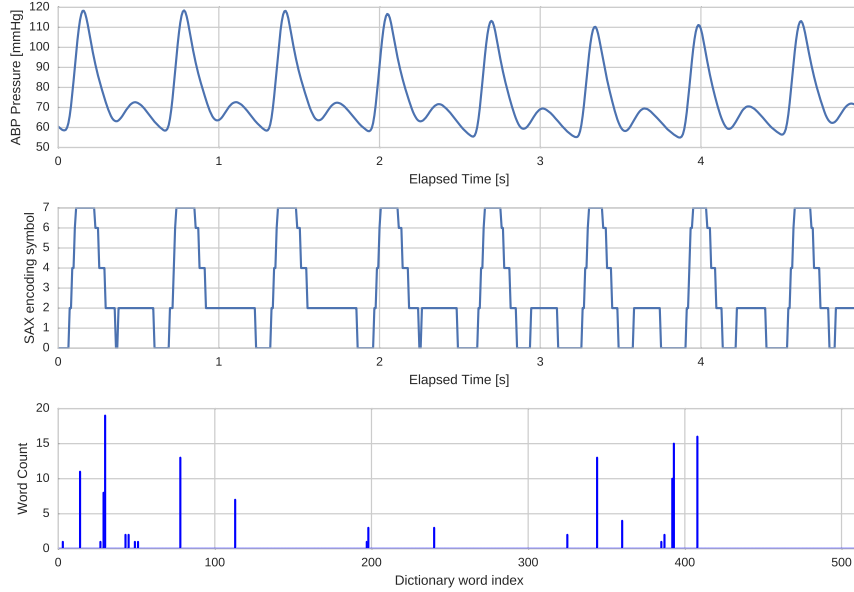
Figure 4.5: Source signal arterial blood pressure (top) SAX encoding (middle) and BoW coding with $\alpha = 3$ and $\omega = 3$ (bottom). The segment length is 5 seconds.

dictionary word length ω is mostly data dependent and can be kept small for smooth time series and should be increased to capture more rapidly changing patterns.

We chose $\alpha = 8$ to capture the pulsatile patterns in ABP and ICP and we chose a word length of size 3 resulting in a total dictionary size of $8^3 = 512$. We did not have to choose an aggregate word length w since we computed the SAX BoW feature on an already down-sampled version of the original signal.

An example showing a 5 second segments of SAX encoded ABP and ICP at the source sample rate of 125 Hz can be seen in Figures 4.5 and 4.6. An example showing 50 second segments of the encoding at sample rate 12.5 Hz can be seen in Figures 4.7 and 4.8.
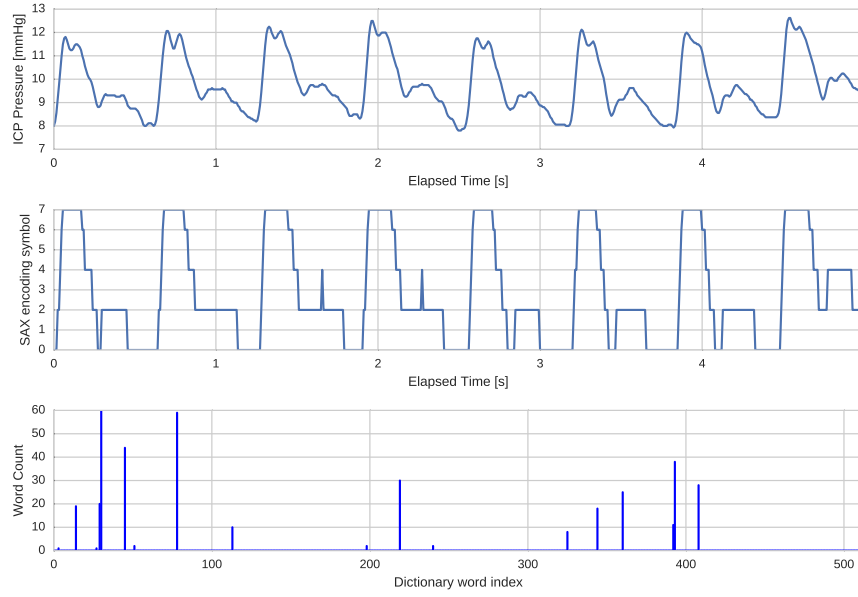
Figure 4.6: Source signal intracranial pressure (top) SAX encoding (middle) and BoW coding with $\alpha = 3$ and $\omega = 3$ (bottom). The segment length is 5 seconds.
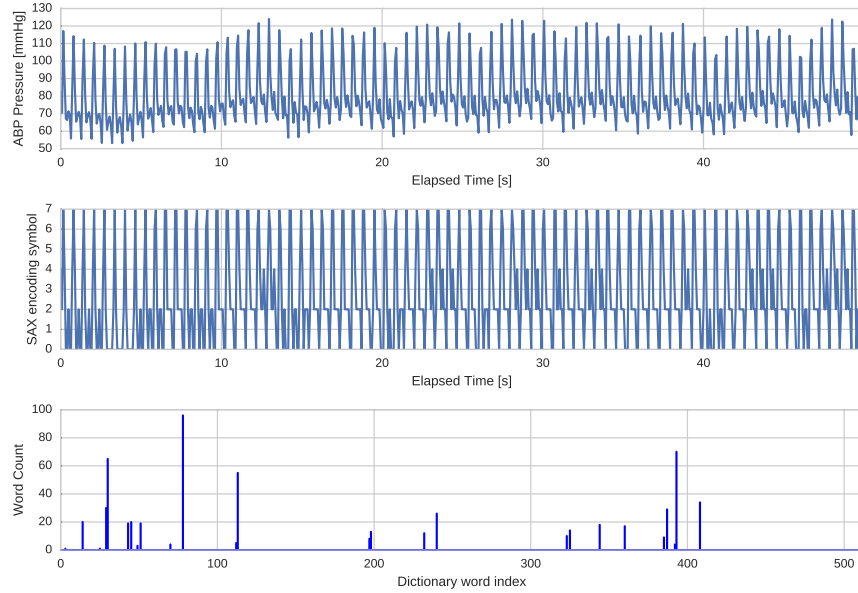
Figure 4.7: Source signal arterial blood pressure (top) SAX encoding (middle) and BoW coding with $\alpha = 3$ and $\omega = 3$ (bottom). The segment length is 50 seconds and the signal has been down-sampled by a factor of 10.
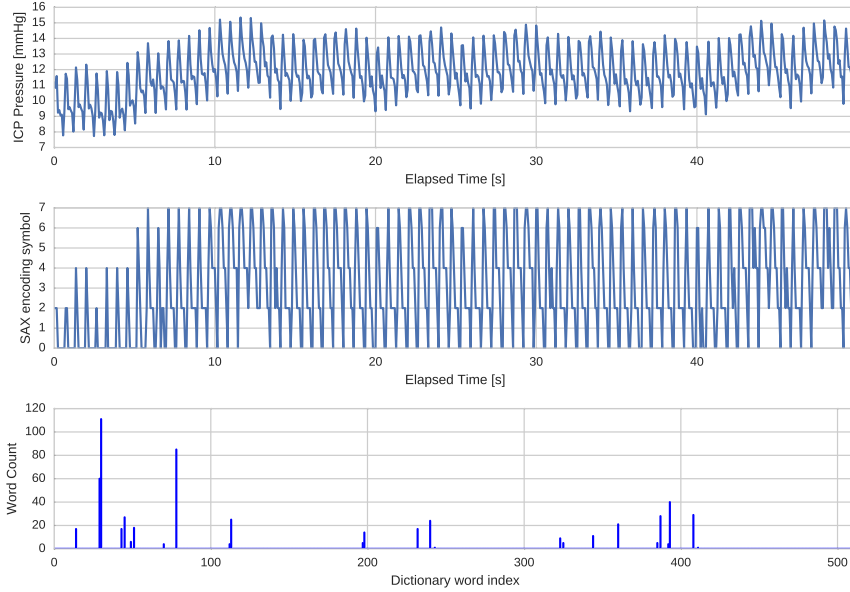
Figure 4.8: Source signal intracranial pressure (top) SAX encoding (middle) and BoW coding with $\alpha = 3$ and $\omega = 3$ (bottom). The segment length is 50 seconds and the signal has been down-sampled by a factor of 10.

4.2.6 Trace

In this work we propose a new feature based on two input vectors x and y of the same length n called $trace_{x,y}$. We interpret the elements x_i and y_i as the coordinates in a two-dimensional space. When looking at the resulting scatter-plot one can expect to see a certain shape. For example, if both input vectors are oscillating with moderate drift, the resulting figure is a circular shape. Figure 4.9 shows a trace of arterial blood pressure (ABP) plotted against intracranial pressure (ICP) where the sample number is encoded in the color of the point. One can clearly see the oscillation of both ABP and ICP and one can also see that both have a smaller shifted sub-oscillation. This sub-oscillation is normally showing as a sub-peak (representing the closing of the aortic valve) in the pulse form and can here be seen in the smaller circle in the middle-left of the figure.

To encode the trace into a feature vector one has to encode the shape of the trace somehow. One option to encode the trace is to encode the angle or the quadrant of the angle of all the lines connecting the points $[x_i, y_i]$ and $[x_{i+1}, y_{i+1}]$. An other possibility is to encode the line length. A third possibility is to overlay the shape with a discrete grid and count the number of points occurring in each grid cell.

We chose the last option because it is most robust to the 'starting point' of

the trace. Since we are tracing ABP against ICP in our specific case, each encoding segment could start at a different position in the pulse. For example, the encoding of two different segments could vastly differ in the angular encoding even if the pulse frequency is the same in both segments when the first sample in the first segment starts at a peak and the first sample in the second segment starts in a valley. If on the other hand we encode by grid coding, the starting point problem diminishes the more samples we add and the more often a full shape is created.

To encode a trace using the discrete grid method we had to solve two problems. First, how to bound the infinitely large plane containing the points $[x_i, y_i]$ and second, how to discretize the plane. Based on the physiological limits of both ABP and ICP we decided to center each trace at its mean and then clip ABP to the range $[-50, 50]$ *mmHg* and ICP to the range $[-10, 10]$ *mmHg*. Furthermore, we discretized the area into a 16×16 grid. This resulted in a feature vector with a reasonable length of 256 values. An example of the encoded ABP-ICP trace shown in Figure 4.9 can be seen in Figure 4.10. The overall shape is still visible but details like the closing of the aortic valve almost disappear.

One important thing to note is that in this feature we lose the location of the shape since we subtract the mean from both input vectors. This information needs to be encoded using a different feature, e.g. in the statistical summaries.
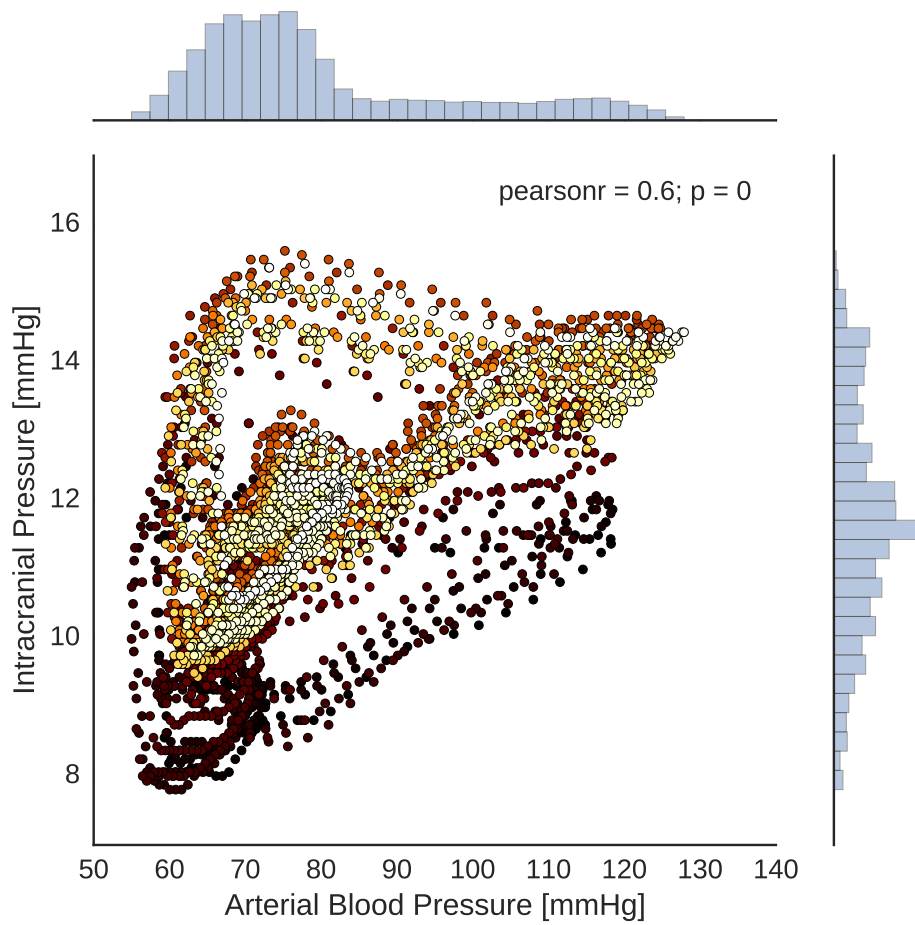
Figure 4.9: Scatter plot of ABP against ICP over a subsegment of 3750 samples (30 seconds). The axes are annotated with a histogram of the respective signal.
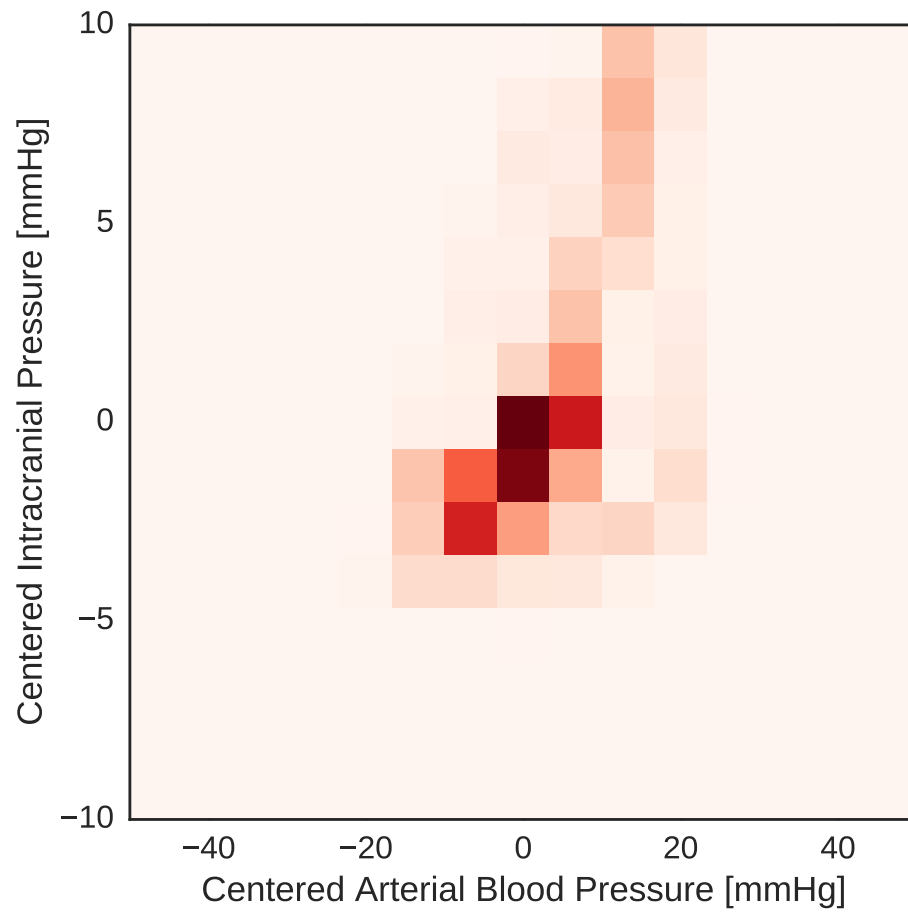
Figure 4.10: Encoding of the scatter plot of ABP against ICP over a subsegment of 3750 samples (30 seconds). We subtracted the mean of both vectors and set the grid clip to -50 to 50 for ABP and to -10 to 10 for ICP.

4.2.7 Wave Morphology

The last feature we propose is a morphology based feature. It segments all the pulses in a given ICP segment and assigns them to a predefined set of ICP pulse classes. We defined those classes by a k-means clustering of all segmented ICP pulses found in the MIMIC II data set. We then chose the number of clusters based on the silhouette score of the segmentation.

When segmenting the pulses we relied on the existing algorithm in the software framework. It segments pulses by first detecting a QRS peak in the ECG signal. Then, it searches for the corresponding peak in the low-pass filtered ICP signal. Finally, it searches left of the peak for the onset of the ICP pulse. After it has detected all onsets it assumes a pulse ends where the next pulse begins and filters out all pulses that have a latency which is bigger than physiologically possible. Details on the implementation of the two segmentation algorithms can be found in the work of Hüser et al. [28].

After we have found the pulse segmentation, we evaluated different ways of encoding the pulses for clustering. Since the feature vector for clustering always needs to have the same length, we need all the pulses to be of the same length. Here we had two options. First, we could resample each pulse to a fixed length (for example 100 samples). This would remove the pulse latency information and could lead to small miss-alignments of peaks but it would retain information on pulse amplitude. Second, we could just align the pulse peaks and the take l samples left of the peak and r samples right of the peak. This would keep latency information but would also have a very high probability of including multiple pulses or not including the full pulse.

We also needed to choose how to encode the resulting feature vector before clustering it. The possibilities were:

- Keep the original pulse
- Subtract the minimum or mean from the pulse
- Normalize the pulse
- Encode the pulse using SAX

Based on empirical evaluation of the different segmentation methods and different encodings we decided to cluster the resampled pulses with their respective minimum removed. We made this decision based on the overall silhouette score of the clustering and the individual silhouette score of each pulse for each cluster. The results of the cluster evaluation can be found in Figure 4.11 and 4.12. We decided to use 20 clusters based on those empirical results. We are aware of the fact that the silhouette score is usually highest for only 3 clusters but we wanted to get a larger variety of pulse shapes to also be able to possibly predict the average form in the future. The plot of the individual silhouette score can be found in Figure 4.13.
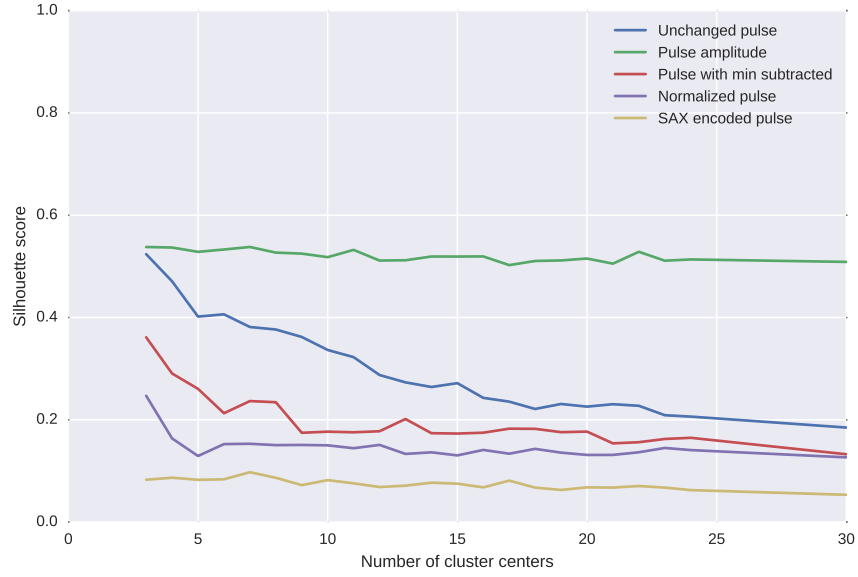
Figure 4.11: Silhouette score for unaligned but resampled ICP pulses. The different colors represent the different ways of encoding the resampled pulses.

Figure 4.14 shows the center shape of each pulse cluster. We can clearly see that pulse amplitude is an important predictor of the pulse class. We can also see that the shape (one, two, or three subpeaks) can be detected. Unfortunately it also shows the current limits of the pulse segmentation method because for example the center at row 3, column 2 clearly shows two consecutive pulses.

Figure 4.15 finally shows the encoding of a short 15 seconds ICP segment. The top graph contains the results of the peak detection routine where the location of each ICP pulse peak is marked by a vertical line. The next graph shows the resampled and concatenated pulses[2]. The second-last graph shows the cluster label of the closest cluster of each segmented pulse. The bottom graph shows the final feature used in our model. It is a frequency count of each cluster label.

---

[2]We concatenated the pulses in this graph to present them in a more compact way. Usually the clustering is done with the individual but resampled pulses.
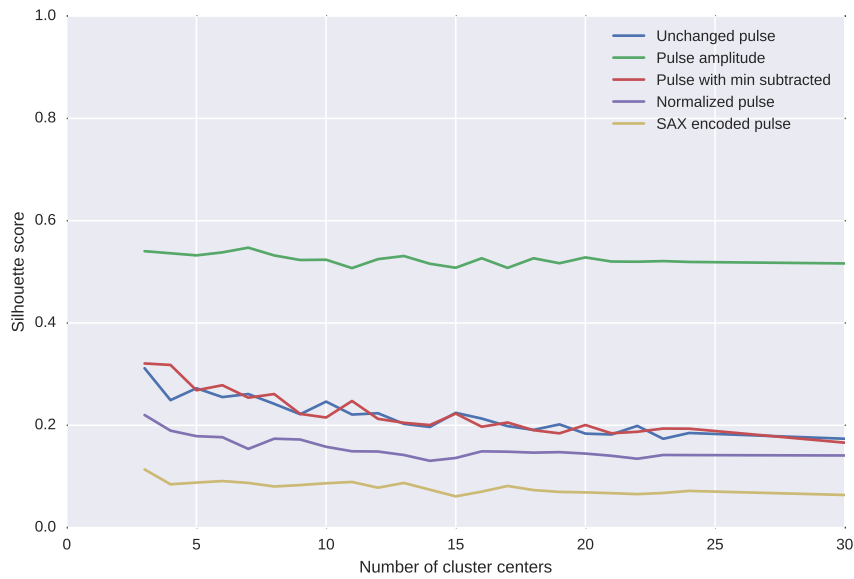
Figure 4.12: Silhouette score for aligned but not resampled ICP pulses. The different colors represent the different ways of encoding the resampled pulses.
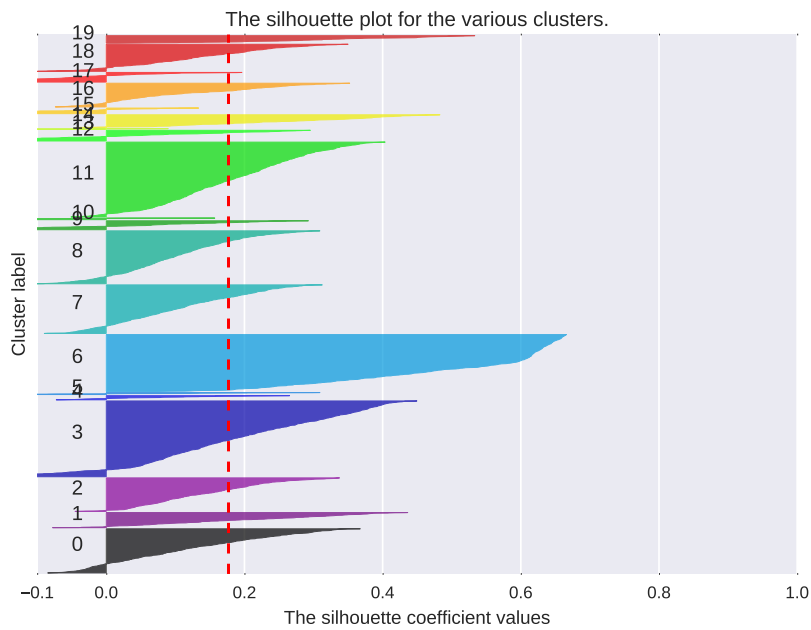
Figure 4.13: Silhouette score the individual pulses when choosing 20 cluster centers on unaligned but resampled pulses without min removed.
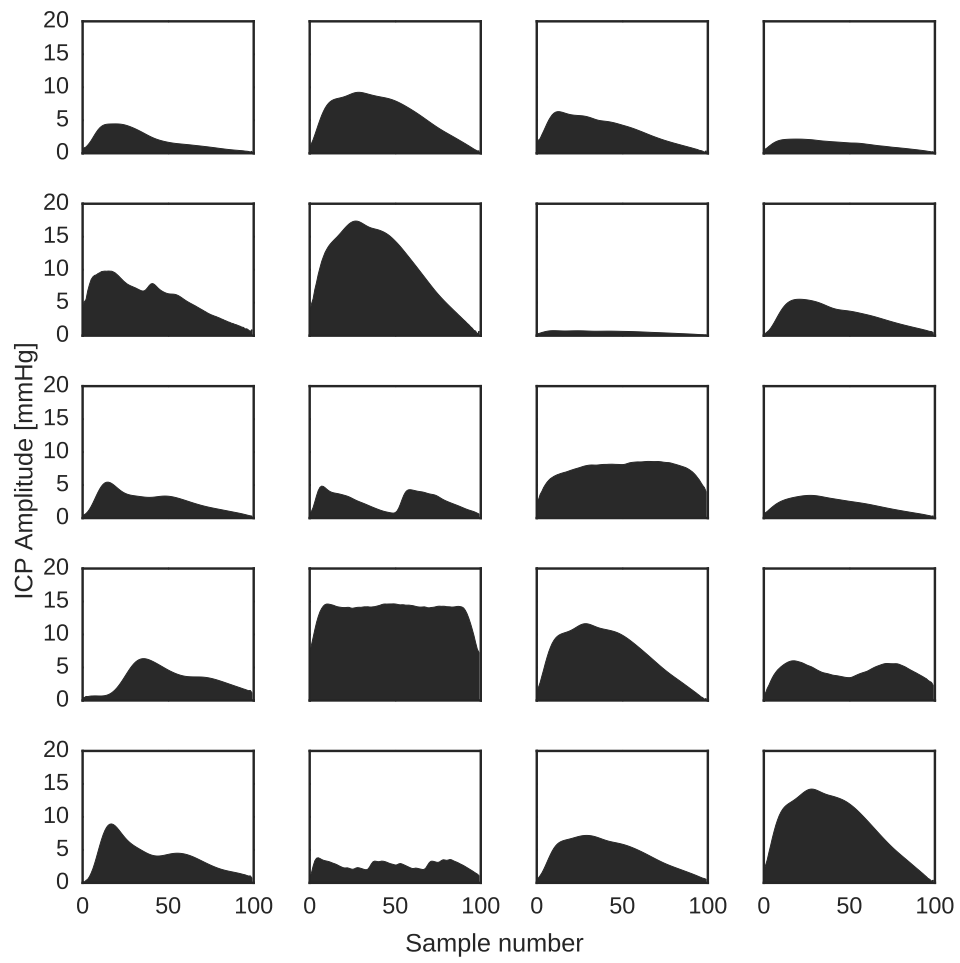
Figure 4.14: The 20 cluster centers with minimum at 0.
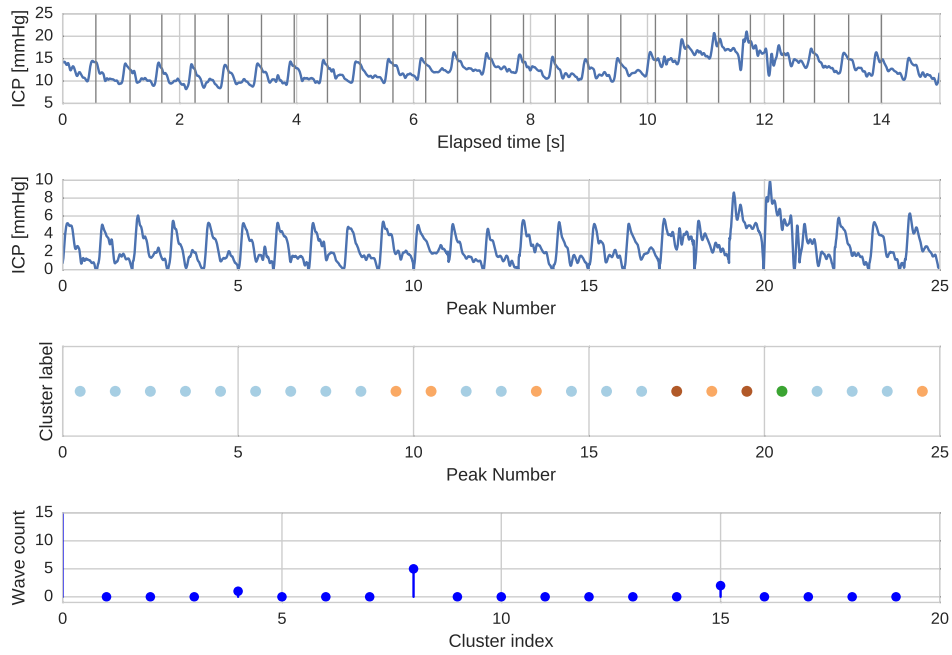
Figure 4.15: Encoding of a 15 second ICP segment by assigning each pulse to its closest cluster. The original signal is at plotted in the top subfigure with the assumed pulse peak marked by a vertical line. The resampled and concatenated pulses can be seen in the upper middle figure. The cluster label is encoded in the color of each point in the lower middle subfigure. And the resulting frequency count feature can be seen in the bottom subfigure.

4.3 Targets

The aim of this work is to predict the future CA capacity. Therefore, the prediction targets considered in this work are:

- autoregulation indices, i.e. PRx, TF, and IAAC, as these cover the different categories of CA indices (correlation based and spectrum based) and are well known in literature.

- monitoring signals, i.e. ABP, ICP, and CPP, as they are key clinical indicators of the patient status. Furthermore, they could be used to compute a subset of the CA indices.

The prediction horizon t ranged from 5 minutes to 2 hours.

4.4 Learning Models

We used linear regression models with an epsilon insensitive loss function regularized using an elastic-net term. We were thus learning the linear function

$$f_{w,b}(x) = w^T x + b$$

where the training error over the feature set X and target vector y is computed as

$$J_{\alpha,\rho,\varepsilon}(w, b) = \frac{1}{n} \sum_{i=1}^{n} L_{\varepsilon}(y_i, f_{w,b}(x_i)) + \alpha R_{\alpha,\rho}(w)$$

where L is the epsilon insensitive absolute loss

$$L_{\varepsilon}(y_i, y_i') = \left\{ \begin{array}{ll} 0, & \text{for } |y_i - y_i'| \leq \varepsilon \\ y_i - y_i', & \text{for } |y_i - y_i'| > \varepsilon \end{array} \right\}$$

and where R is the elastic-net regularization term

$$R_{\alpha,\rho}(w) = \frac{\rho}{2}||w||_1 + \frac{1-\rho}{2}||w||_2^2$$

Notice, that we can append a constant column of $\mathbf{1}s$ to the feature matrix X to be able to append the bias b to the weight vector w. We can thus omit the variable b in further formulas.

For each feature set we trained the model by minimizing the error $J_{\alpha,\rho,\varepsilon}(w)$ on the training set X_{train}, y_{train} as $\text{argmin}_w J_{\alpha,\rho,\varepsilon}(w)$ using stochastic gradient

descent (SGD). We picked SGD because it is able to handle data sets of 500'000 rows with 7'000 feature columns.

SGD approximates the true gradient of the objective function $J(w)$ by the gradient at a single example

$$w^{(t+1)} = w^{(t)} - \eta^{(t+1)} \nabla J_i(w)$$

In our case, the weight update becomes

$$w^{(t+1)} = w^{(t)} - \eta^{(t+1)} \nabla \left[L_\varepsilon(y_i, f_w(x_i)) + \alpha R_{\alpha,\rho}(w) \right] \tag{4.1}$$

We adjust the step size η given an initial learning rate t_0 as described in [6, 52]:

$$\eta^{(t)} = \frac{1}{\alpha(t_0 + t)}$$

In every epoch SGD iterates through a random permutation of the training set. After each epoch we compute a validation error on a separate set X_{val}. We stop when the validation error has not decreased for k rounds.

4.5 Software Framework

The details on the initial implementation of the framework can be taken from the work of Hüser et al. [28]. Here we summarize the overall structure.

The initial data set is assumed to be stored inside an arbitrary data store. The feature and target construction programs then work on a per-record basis. For feature construction, the framework first loads all required signals of a record from the data store, then it splits each signal up into windows of 10 seconds, next it preprocesses the windows to remove artifacts and mark windows with missing data or noise as invalid. Then, it resamples each window to different sampling rates required for feature construction. Finally, it constructs the configured features and stores the resulting features back into a data store, where each is annotated with time stamps and the record identifier.

For target construction the framework first loads all features needed for target construction into memory and then applies the target construction functions to the full feature matrix.

Next we will take a more detailed look at the enhanced software framework. The principal concepts of the original software framework have been kept

the same. These four main concepts were: Online computation of features, multi-scale history of signals, caching of constructed features, and a separation of the overall process into a preprocessing, a feature construction, and a target construction phase.

4.5.1 Online Computation of Features

This is the fundamental model embedded in the software framework. It is an important design decision because it forces possible feature construction algorithms to also be applicable to real world scenarios where signals will undoubtedly arrive in a streaming fashion.

4.5.2 Multi-Scale History

An other important concept is the multi-scale history. It provides a view on the input data on different sampling rates, thus allowing the analysis of the input signal at different time scales. It works by keeping a history of past input windows for each signal. Whenever a new window is added to the multi-scale history, the original samples are resampled to each level of the multi-scale history and prepended to the buffer storing the samples of the corresponding signal and sampling rate. Each level has a maximum history length for which it must retain its samples. When new samples are added even though the buffer contains samples up until the maximum history length, the oldest values are discarded to make room for the new.

The framework can query each input signal at any defined sampling rate for any history length up to the maximum history length. If the buffer does not contain sufficient samples to cover the requested time frame, it returns a segment marked as invalid.

4.5.3 Caching of Constructed Features

To reduce processing time the original framework introduced caching of constructed features. We extended the way how features could depend on other features and have retained the way features are cached.

4.5.4 Pipeline Architecture

In the original and in the current version of the software framework it is only possible to create targets from features. This is due to the fundamental difference in the approach of how the two are computed. Feature construction is modeled as an online process, where batched up windows of input samples are processed and stored and where results that are not needed anymore are discarded.

Target construction on the other hand is modeled as an offline process. Target construction functions have access to the full feature set at once. Therefore, they are functions designed to do relatively simple operations and not use too much memory since the feature set could potentially be very large. Examples for target construction are: Shifting the values back to create a predictive target, normalizing columns or rows, or converting from continuous values to binary values by thresholding.

After giving a general overview on the feature and target construction process we will now list the most important enhancements made to the original software framework.

4.5.5 Enhancements

Configuration   In the original framework, each feature had a unique descriptor string consisting of the construction function's name and the number and names of its arguments. When configuring the feature and target creation process one had to write the feature descriptors for each feature into a configuration file. This approach was rather fragile since the final set of features contained about 1'000 feature descriptors and it was therefore very hard to check if all the descriptors had the correct arguments at the correct scale and history length.

We improved the framework by moving all configuration into code. The unique feature descriptor is now constructed automatically from a feature descriptor object (FDO) and only used for storage. An FDO must be created inside a Python script and has the same features as the original string based descriptor. Furthermore, we also added an input descriptor object (IDO) to describe raw signal data.

As a result can FDOs now be checked automatically and configuration errors can quickly be detected. Also, when working inside an integrated development environment one is able to get feedback on errors made when creating a FDO. FDOs can now also be generated programmatically such that the full feature set can be created in a more concise and readable way. Since a feature set is then a list of FDOs one can use the standard Python tools to manipulate the list, for example to remove features containing a specific input signal. Later in the modeling process it is also possible to directly reference features that have been created in earlier script by simply importing the feature creation script. Thus one can be sure that one is always working with the correct set of features.

Feature Construction Function Arguments   In the original version of the framework it was only possible to call feature constructors with the preprocessed signals or numerics. The feature constructor functions themselves

then declared their dependency on other feature construction functions. We inverted this process such that the user is required to explicitly define each argument to a feature construction function.

This thus allows the user to arbitrarily nest FDOs by providing a mix of one or more FDOs and IDOs as arguments to an other FDO. Like in the original version of the framework, the results of each feature construction function are cached for the current window to speed up feature construction. This is mainly noticeable when many features depend on the same intermediate feature. To compute the features we traverse the dependency tree of each FDO in a depth first manner. Allowing arbitrary nesting of FDOs led to better modularization of feature construction code and sped up feature construction in general. On the other hand the user now needs to be more careful when constructing features since he is required to give the correct intermediate feature as an argument to the final feature and argument checks are limited. This approach could definitely be improved by providing convenience functions constructing the correct feature dependency tree for the user.

We also experimented with parallelizing feature construction since most features could be constructed independently and we only had to synchronize access to the cache storing intermediate features construction results. However, it turned out that the locking overhead was too big and we also did not achieve good enough parallelization. This is due to the global interpreter lock (GIL) in python which allows only one Python thread to be executed at any time. Even though we did many computations in NumPy, which releases the GIL, we still did not achieve a speed up and thus removed parallelization on a per-feature basis and instead rely on parallelization on a per-record basis.

HDF 5 Based Storage   We switched from CSV to HDF 5 to store the initial data set from Mimic and Cambridge. This decreased record loading time due to the fact that records stored in HDF 5 files can be loaded directly into memory as NumPy arrays, skipping the expensive text parsing step of CSV files.

Configurable Sample Rates   We extended the software framework to be able to handle different sample rates. This means that the sample rate of waveform signals and numeric signals can be set independently at the start of a feature construction run. In the original framework these two values were hardcoded and not adjustable since parts of the framework assumed them to be hardcoded.

Configurable Window Size   We extended the software framework to be able to handle different window sizes. This means that at the start of the feature

construction run one can specify the number of samples in the window that is added to the multi-scale history in each step. In the original framework this value was hardcoded to 30 seconds and not adjustable.

Arbitrary Scales In the original software framework the sampling rate of the different levels of the multi-scale history were fixed in the framework's source code. They could be adjusted since the code directly referred to them but this was not very user friendly.

We extended the framework by adding an automatic multi-scale history construction step. This step analyses the FDO forest and computes all the required sampling rates and history lengths for the input signals. As a byproduct we now only load signals into the multi-scale history that are later required for the construction of a feature and we also set the maximum history length of each signal to the longest history required by any FDO.

We also added a check to verify that the down-sampling rate of each level of the multi-scale history is an integer factor of the original sampling rate. Otherwise the actual sample rate of a level would get truncated to the next lower integer factor.

Speed Improvement We enhanced the framework by moving expensive feature construction function from Python code into JIT compiled C code by using the NUMBA and the Cython framework. This allowed us to decrease the runtime of the feature construction step. Furthermore, since we are now more explicitly making use of optimized functions in NumPy, the underlying linear algebra library can autoparallelize some computations[3].

4.6 Library Dependencies

The original software framework already relied on different Python libraries to implement parts of preprocessing, feature-, and target construction. Here we summarize the extended list of Python libraries that we used in the extended framework. The biggest dependencies are the two libraries NumPy and SciPy for numerical computation and ScikitLearn for machine learning. Numba, Cython, and Bottleneck were used for performance improvements and PyWavelets and MLpy were used for small parts of feature construction. Next follows the list of dependencies with a small description of each library.

NumPy an N-dimensional array and linear algebra package (http://www.numpy.org/).

---

[3]Results on the achieved performance gains can be found in Figure A.9 in the appendix.

SciPy provides many user-friendly and efficient numerical routines such as routines for numerical integration and optimization (`https://scipy.org/scipylib/index.html`).

ScikitLearn a machine learning library covering preprocessing, modeling, and evaluation (`http://scikit-learn.org/stable/`).

Numba generates optimized machine code from python code using the LLVM compiler infrastructure at import time, runtime, or statically (`http://numba.pydata.org/`).

Cython an optimising static compiler for both the Python programming language and the extended Cython programming language (`http://cython.org/`).

Bottleneck a collection of fast NumPy array functions written in Cython (`http://berkeleyanalytics.com/bottleneck/`).

PyWavelets a wavelet transform software for Python (`http://www.pybytes.com/pywavelets/`).

MLpy a Python module for Machine Learning built on top of NumPy/SciPy and the GNU Scientific Libraries (`http://mlpy.sourceforge.net/`).

Finally we need to mention the two small software kits, pyeeg (`https://code.google.com/p/pyeeg`) and mne-tools (`https://github.com/mne-tools/mne-python`) from which we and the original framework take some feature creation algorithms.

Learning Algorithm

The only algorithm available in SciKit-Learn for our criteria is stochastic gradient descent (SGD). We also evaluated the framework Keras [4] which is usually used for training deep learning architectures and built on top of Theano [5]. However, the models created using the Keras framework took longer to train and were only able to optimize absolute error without an epsilon insensitive region.

We wrapped the SGD algorithm implemented in the SciKit-Learn framework to implement an early stopping criterion. For this we split off a small validation set from the training set and evaluated the model on this validation set after each epoch, i.e. after each full iteration through the training set. Using an early stopping mechanism allowed us to reduce training time by stopping after a model's accuracy did not improve anymore.

---

[4] `http://keras.io/`
[5] `http://deeplearning.net/software/theano/`

4.6.1 Feature Set Abstraction

The FeatureSet implementation differentiates between four different subsets of features: non-normalized and normalized features and sparse or dense features. All features added as normalized features had their mean subtracted and were divided by their standard deviation. Non-normalized were left as is. When sparse features were added we had to load the data set in a different way to not use too much memory. However, loading dense features is faster than loading sparse features and thus we generally prefer it.

When loading the feature matrix into memory using the FeatureSet, the user is also able to specify the data type of the feature matrix and if the normalization steps should already be applied. If the user chooses not to apply the transformation immediately after loading the data, the FeatureSet builder returns the normalization pipeline. The pipeline can then be used to apply the transformations at a later point in time.

Deferring the normalization step also allows the user to further customize the transformation pipeline and finally append a learning model. Since the full pipeline is also just a model, it can be trained as a whole and then be serialized to disk. This allows others to easily reproduce all results and reevaluate the model after it has been trained. Also, it allows the user to continue training the model at a later point in time.

The user is also able to specify an individual feature preprocessing pipeline for dense non-normalized, dense normalized, and sparse features. This for example allows the user to apply an additional logarithmic or exponential transformation on one of the feature sets before they get further processed (imputation and normalization happen after a user's custom pipeline). Finally, the user is able to specify a threshold percentage to drop rows or columns containing more missing values than the given threshold.

4.6.2 Handling of Missing Values

We evaluated different methods to handle missing values in addition to the approaches already proposed by Hüser et al. [28]. We tried to drop columns with many or any missing values and we tried to drop rows with many or any missing values. Finally, we settled with Hüser's approach to replace missing values with zero since no method seemed to considerably improve the resulting accuracy of the model.

4.6.3 Normalization

We normalized feature columns as proposed by Hüser et al. [28] by normalizing each column based on existing values. This means that we normalize while the data still contains missing values but we ignore missing values

during computation of the normalization parameters. Since we thus subtract the mean from every column, missing values are later replaced by each column's mean value. Non-normalized columns should therefore be zero mean or not contain any missing values.

4.6.4 Feature Selection

We also tried to drop columns according to univariate statistical tests. We dropped columns having a p-value larger than different thresholds but could not increase model accuracy in any case. Thus we ignored this step in the final experiments.

Chapter 5

Evaluation and Results

In this chapter we will explain the experiments we ran and discuss their results. We designed the experiments to evaluate the clinical applicability of our machine learning models.

We will first describe the experimental design composed of our feature selection, target selection, and prediction horizon selection. Then we will describe the training and our evaluation procedure. Finally, we will analyze the results of the experiments we ran and compare them to results reported by related work.

5.1 Experimental Design

To evaluate the predictive power of our proposed machine learning model we compared different feature sets at different prediction horizons for each prediction target. We also added a baseline which assumes that the prediction target does not change from the current value.

We compared the selected baseline with a baseline predicting the average target and found no profound difference between them. Table 5.1 reports the difference in MAE of between the two baselines. A negative difference indicates that the selected baseline was better. We see that the average-predicting baseline only has a marginally better prediction accuracy. Thus, we selected the zero-predicting baseline.

Target	30 min	60 min	120 min
ABP	0.15 ± 0.0	0.22 ± 0.0	0.33 ± 0.0
CPP	0.15 ± 0.0	0.16 ± 0.0	0.21 ± 0.0
IAAC	0.0 ± 0.0	0.0 ± 0.0	0.002 ± 0.0
ICP	-0.001 ± 0.0	0.003 ± 0.0	0.032 ± 0.0
PRx	0.001 ± 0.0	0.003 ± 0.0	0.001 ± 0.0
TF	0.0 ± 0.0	0.001 ± 0.0	-0.0 ± 0.0

5.1.1 Feature Sets

We divided the full set of features into 9 subsets. Each subset is focused on a specific class of features. This way we tried to evaluate the predictive power of each feature set in isolation for each prediction target and horizon.

An alternative to splitting the features up into subsets would be to recursively add or eliminate features from the full feature set until the best accuracy is achieved. Also, one could use the magnitude of the regularizer's coefficients used in the linear model to select a set of highly predictive features. However, doing this for all the prediction targets at all different prediction horizons turned out to take too long, especially since the full feature set contains almost 7000 feature columns from about 1000 distinct features.

We now list the 9 selected feature sets plus an additional reference set. Each features set also had the current value of the target to be predicted as a feature column.

History of target contain the last 5 minutes of the target value, i.e. the last 50 30-second mean values or the last 50 CA indices. It is a simple reference feature set.

Statistical summaries contain the values x_{min}, x_{mean}, x_{median}, x_{max}, x_{std}, x_{var}, x_{kurt}, x_{skew}, x_{norm}, x_{slope} for the windows with length 30 seconds (125 Hz / 1 Hz), 5 minutes (12.5 Hz / 0.1 Hz), and 25 minutes (1.5 Hz / 0.1 Hz) of the signals Heart Rate (HR), Breathing Frequency (RESP), Partial O_2 pressure (SpO2), Intracranial Pressure (ICP), Arterial Blood Pressure (ABP), and Cerebral Perfusion Pressure (CPP). See Section 4.2.1 for a more detailed description of the values.

Extended Statistical summaries extend the statistical summaries by adding the logarithm, exponential, and square root transform of each value.

Frequency features extend the statistical summaries and contain the Fourier transform and the Wavelet transform features as described in Section 4.2.2 and 4.2.3.

CA indices extend the statistical summaries and contain x_{PRx}, x_{RAP}, x_{PAx}, x_{TF}, x_{SLOW}, and x_{IAAC} computed over 10 minute segments. They also contain the trend of each of those indices computed over 20 minute segments via the least squares on the index values computed from a moving window of 10 minutes overlapping in 9 minutes. See Section 4.2.4 for a more detailed description of the indices.

Extended CA indices extend the CA indices by adding the same logarithmic, exponential, and square root transform of all values as described in Extended Statistical summaries.

Huser et al. contains all the features proposed by Hüser et al. [28]. For details see the reference work.

ICP pulse morphology extend the statistical summaries and contain the normalized frequency count of each cluster index. See Section 4.2.7 for a more detailed description of the values.

SAX BoW extend the statistical summaries and contain the bag of word frequency counts of the SAX encoding of ABP, CPP, and ICP for windows of length 5 minutes (12.5 Hz) and 25 minutes (1.5 Hz). See Section 4.2.5 for a more detailed description of the values.

ICP-ABP trace extend the statistical summaries with the encoded ICP-ABP trace shape. See Section 4.2.6 for a more detailed description of the values.

5.1.2 Prediction Horizons

We selected the prediction horizons $T = [5, 10, 20, 30, 60, 90, 120]$ minutes. A prediction horizon below 30 minutes is likely not clinically relevant but interesting from an evaluation point of view.

5.1.3 Prediction Targets

We selected intracranial pressure (ICP), arterial blood pressure (ABP) cerebral perfusion pressure (CPP), pressure reactivity index (PRx), Single Wave ICP-ABP Amplitude Correlation (IAAC), and transfer function index (TF) as prediction targets. We selected these physiological signals because they are clinically relevant and usually monitored in intensive care. Additionally, one would be able to compute the PRx based on the predicted physiological signals[1]. We selected these CA indices because they cover the different categories of CA indices well.

In our forecasting model we then predicted the change in the target, i.e. the difference between the current value of the target and the value of the target t minutes in the future. We assume that this is a lot easier for a linear model to predict. If we would want to predict the absolute value instead, we could simply add the current non-normalized target value as feature and the model would then need to learn to use it.

In Table 5.1 we summarized the statistics of each target [2]. It is interesting that all prediction targets have zero-mean or almost zero-mean difference from the current value to the value at horizon t. However, the standard deviation increases over time. This also indicates why the zero-predictor is

---

[1]We omitted the results of this comparison in this work and leave it for future work.

[2]We summarize only MIMIC II data but have observed the same results on the Cambridge data set.

Horizon	ABP	CPP	ICP
5 min	-0.01 ± 7.97	-0.01 ± 8.20	-0.00 ± 3.20
10 min	-0.03 ± 9.00	0.00 ± 9.17	-0.01 ± 3.48
20 min	-0.06 ± 10.12	-0.03 ± 10.33	-0.01 ± 3.87
30 min	-0.08 ± 11.04	-0.06 ± 11.22	0.00 ± 4.18
60 min	-0.08 ± 12.55	-0.05 ± 12.76	0.01 ± 4.71
90 min	-0.10 ± 13.21	-0.07 ± 13.44	0.00 ± 4.98
120 min	-0.12 ± 13.50	-0.06 ± 13.68	-0.00 ± 5.08
Horizon	PRx	IAAC	TF
5 min	0.00 ± 0.33	0.00 ± 0.20	-0.00 ± 0.06
10 min	-0.00 ± 0.40	-0.00 ± 0.25	-0.00 ± 0.07
20 min	-0.00 ± 0.42	-0.00 ± 0.27	-0.00 ± 0.08
30 min	-0.00 ± 0.43	-0.00 ± 0.27	-0.00 ± 0.08
60 min	-0.00 ± 0.44	-0.00 ± 0.28	-0.00 ± 0.08
90 min	-0.00 ± 0.45	-0.00 ± 0.29	0.00 ± 0.08
120 min	-0.00 ± 0.45	-0.00 ± 0.29	0.00 ± 0.08

Table 5.1: Descriptive statistics on all predicted targets. For all targets we predicted the change from the current value to the value at the prediction horizon. Thus a zero-mean targets does not change on average.

a good baseline. It simply does predict the most likely target but it will have increasing error with increasing variance.

5.1.4 Model Evaluation

To evaluate a learning model we did leave-one-patient-out cross-validation. For the MIMIC II data set this resulted in 26 individual test scores, for the Cambridge data set this resulted in 5 individual test scores. To compute the overall score we took the macro average of the individual scores to weight each patient the same. Otherwise, patients with longer recording sessions would bias the overall result. We chose leave-one-patient-out cross-validation because it seemed to be the hardest problem and modeled the real world scenario the best, where the algorithm also does not have access to the patient's future recordings.

Since the MIMIC II data set has some really long recordings (up to 7 days), we subsampled the data for development. For this we took a maximum of 2500 rows per patient but retained the leave-one-patient-out cross-validation scheme.

5.1.5 Hyperparameter Search

The chosen model and the optimization algorithms have hyperparameters. After some initial evaluation and following recommendations by Bengio [5] and by Bottou [6] we chose to optimize the initial learning rate η_0, the regularization coefficient α, the l1-ratio ρ of Elastic-Net, and the size of the epsilon insensitive region ε.

For each of these parameters we defined a statistical distribution together with its parameters based on our assumption on how the parameters are distributed. We then iteratively drew samples from the hyperparameter distribution and did a three-fold cross validation on the full data set to find the best parameter setting. After the first run we were able to extract the best performing hyperparameter configuration of each model and to adjust the hyperparameter-distributions for the next run.

5.2 MIMIC II

For the MIMIC II data set we have the following experimental setup:

$$\begin{aligned}
\text{Horizons } T &= \{5, 10, 20, 30, 60, 90, 120\} \\
\text{Feature Sets } F &= \{history, stats, frequency, trace, index, sax, huser, morph\} \\
\text{Targets } P &= \{ICP, ABP, CPP, PRx, IAAC, TF\} \\
\text{Records } R &= \{3106263, 3142868, 3148126, 3160820, 3169632, 3189000, \\
&\quad 3270980, 3309132, 3319401, 3365681, 3367596, 3487247, \\
&\quad 3543187, 3516004, 3562822, 3624651, 3629298, 3642023, \\
&\quad 3655233, 3656395, 3668415, 3688532, 3693937, 3700665, \\
&\quad 3774557, 3938777\}
\end{aligned}$$

In the experiments we then trained a model for each combination of horizon, feature set, and target. The training steps involves the leave-one-patient-out cross-validation scheme and the hyperparameter search described in Section 5.1.4 based on the mean absolute error (MAE) (To show an insight on the results of hyperparameter selection we refer to Figure A.11 in the appendix). We did the leave-one-patient-out cross-validation on the subset of 26 records containing ICP recordings. For both the 3-fold cross-validated hyperparameter search models and the leave-one-patient-out cross-validated final models we store the mean absolute error and the standard deviation of the cross validation for later analysis. More precisely, the evaluation of each model looks the following:

The mean absolute error of each hyperparameter search cross-validation is stored in the array $HP_{t,p,f,h}$, the corresponding standard deviation in

$HP\_STD_{t,p,f,h}$, both of size $|T| \times |P| \times |F| \times$ #hyperparameter_search_iterations. The mean absolute error of the leave-one-patient-out cross-validation is stored in the array $GR_{t,p,f,r}$ of size $|T| \times |P| \times |F| \times |R|$.

Algorithm 5.1: Model evaluation.

```
1  input: set T, set F, set P, set R
2  output: HP_{t,p,f,h}, HP_STD_{t,p,f,h}, GR_{t,p,f,r}
3  begin
4    foreach horizon t in T:
5      foreach target p in P:
6        foreach featureset f in F:
7          foreach set of randomly drawn hyperparameters h:
8            compute HP_{t,p,f,h} and HP_STD_{t,p,f,h} using 3−fold cross−validation
9          end
10
11         create new model m from the hyperparameters argmin_h HP_{t,p,f,h}
12
13         foreach record r in R:
14           train m on data R \ {r}
15           compute GR_{t,p,f,r} by testing m on r
16         end
17       end
18     end
19   end
20 end
```

Based on the resulting array GR we compute the models performance using the macro average over all records R, i.e. $G_{t,p,f} = \frac{1}{|R|} \sum_r^R GR_{t,p,f,r}$, and the corresponding standard deviation $G\_STD_{t,p,f}$.

We find the best model for each target and horizon as $BEST_{t,p} = \text{argmin}_{f \in F} G_{t,p,f}$, its error as $BEST\_ERR_{t,p} = G_{t,p,BEST_{t,p}}$ and its standard deviation $BEST\_STD_{t,p} = G\_STD_{t,p,BEST_{t,p}}$.

We proceeded in the same way to determine the best model when using the root mean squared error (RMSE)[3].

We chose the MAE to determine the accuracy of the model when the model should be good in the average case but must not focus on extreme events. The other way around we chose the RMSE to report the model accuracy putting more weight on the extreme cases. We take the root of the mean squared error to make them easier to interpret reporting the error in the metric of the prediction target.

In the following sections we report the MAE and the RMSE at the different prediction horizons for each prediction target. All errors are reported as mean $\pm$ standard deviation.

---

[3]Please keep in mind that the models were trained with the epsilon-insensitive absolute loss. Thus a higher accuracy is most likely achievable if the models would be trained with a squared loss.

Feature Set	5 min	30 min	60 min	120 min
Baseline	1.42 ± 0.67	2.15 ± 1.01	2.45 ± 1.22	2.66 ± 1.36
Statistical summaries	1.42 ± 0.67	2.14 ± 1.01	2.40 ± 1.18	2.61 ± 1.30
Ext. stat. summaries	1.42 ± 0.67	2.06 ± 0.95	2.36 ± 1.15	2.57 ± 1.21
Frequency features	1.42 ± 0.67	2.08 ± 0.98	2.37 ± 1.17	2.57 ± 1.34
CA indices	1.42 ± 0.67	2.15 ± 1.00	2.45 ± 1.20	2.61 ± 1.30
Ext. CA indices	1.42 ± 0.67	2.15 ± 1.01	2.39 ± 1.18	2.58 ± 1.28
Huser et al.	$\underline{1.36 \pm 0.64}$	2.01 ± 0.94	$\mathbf{2.24 \pm 1.08}$	$\mathbf{2.41 \pm 1.18}$
ICP pulse morphology	1.42 ± 0.67	2.10 ± 0.97	2.37 ± 1.18	2.58 ± 1.28
History of target	$\mathbf{1.35 \pm 0.62}$	$\mathbf{1.94 \pm 0.94}$	$\underline{2.36 \pm 1.18}$	$\underline{2.47 \pm 1.23}$
SAX BoW	1.42 ± 0.67	14.70 ± 58.12	2.45 ± 1.22	2.66 ± 1.36
ICP-ABP trace	1.42 ± 0.66	2.15 ± 0.99	2.45 ± 1.19	2.66 ± 1.33

Figure 5.1: Average mean absolute error (MAE) and standard deviation of all models and the baseline model when predicting Intracranial Pressure (ICP).

Prediction Accuracy Over Time

We have summarized the results of our experiments in Figure 5.5 for ABP, in Figure 5.6 for CPP, in Figure 5.8 for ICP, in Figure 5.7 for IAAC, in Figure 5.9 for PRx, and in Figure 5.10 for TF.

Each figure starts with two graphs showing how the MAE and the RMSE increases for an increasing prediction horizon. For each prediction target there also exists a summary table listing the prediction accuracy. We show the results for ICP in Table 5.1 and 5.2 and for PRx in Table 5.3 and 5.4. Tables for the remaining prediction targets can be found in the appendix.

Feature Set	5 min	30 min	60 min	120 min
Baseline	2.35 ± 1.14	3.19 ± 1.46	3.51 ± 1.66	3.68 ± 1.79
Statistical summaries	2.35 ± 1.14	3.17 ± 1.45	3.40 ± 1.60	3.49 ± 1.67
Ext. stat. summaries	2.35 ± 1.14	2.95 ± 1.31	5.31 ± 10.31	3.40 ± 1.59
Frequency features	2.35 ± 1.14	3.05 ± 1.39	3.36 ± 1.58	3.44 ± 1.72
CA indices	2.35 ± 1.14	3.19 ± 1.44	3.51 ± 1.63	3.56 ± 1.71
Ext. CA indices	2.35 ± 1.14	3.19 ± 1.46	3.40 ± 1.62	3.48 ± 1.69
Huser et al.	<u>2.22 ± 1.06</u>	<u>2.94 ± 1.33</u>	**3.16 ± 1.45**	<u>3.29 ± 1.53</u>
ICP pulse morphology	2.35 ± 1.14	3.11 ± 1.37	<u>3.32 ± 1.57</u>	3.52 ± 1.68
History of target	**2.18 ± 1.03**	**2.80 ± 1.29**	3.33 ± 1.58	**3.28 ± 1.59**
SAX BoW	2.35 ± 1.14	20.66 ± 81.47	3.51 ± 1.66	3.68 ± 1.79
ICP-ABP trace	2.35 ± 1.13	3.17 ± 1.44	3.47 ± 1.63	3.65 ± 1.77

Figure 5.2: Average root mean squared error (RMSE) and standard deviation of all models and the baseline model when predicting Intracranial Pressure (ICP).

Feature Set	5 min	30 min	60 min	120 min
Baseline	0.25 ± 0.05	0.34 ± 0.06	0.35 ± 0.06	0.36 ± 0.07
Statistical summaries	0.23 ± 0.04	<u>0.29 ± 0.05</u>	0.29 ± 0.05	0.29 ± 0.06
Ext. stat. summaries	0.23 ± 0.04	0.29 ± 0.05	0.29 ± 0.05	0.30 ± 0.06
Frequency features	0.23 ± 0.04	0.29 ± 0.05	<u>0.29 ± 0.05</u>	0.29 ± 0.06
CA indices	<u>0.23 ± 0.04</u>	**0.28 ± 0.05**	0.32 ± 0.26	0.29 ± 0.06
Ext. CA indices	**0.22 ± 0.04**	0.29 ± 0.05	**0.29 ± 0.05**	**0.29 ± 0.05**
Huser et al.	2672.61 ± 9253.99	0.31 ± 0.05	0.32 ± 0.05	0.36 ± 0.06
ICP pulse morphology	0.23 ± 0.04	0.29 ± 0.05	0.29 ± 0.05	<u>0.29 ± 0.05</u>
History of target	0.23 ± 0.04	0.29 ± 0.05	0.29 ± 0.06	0.29 ± 0.06
SAX BoW	0.25 ± 0.05	0.30 ± 0.05	0.34 ± 0.12	0.32 ± 0.06
ICP-ABP trace	0.23 ± 0.04	0.30 ± 0.05	0.31 ± 0.06	0.32 ± 0.06

Figure 5.3: Average mean absolute error (MAE) and standard deviation of all models and the baseline model when predicting Pressure Reactivity Index (PRx).

Feature Set	5 min	30 min	60 min	120 min
Baseline	0.33 ± 0.06	0.44 ± 0.07	0.45 ± 0.08	0.45 ± 0.08
Statistical summaries	0.30 ± 0.05	$\underline{0.35 \pm 0.06}$	0.36 ± 0.06	0.36 ± 0.06
Ext. stat. summaries	0.29 ± 0.05	0.36 ± 0.05	$\underline{0.36 \pm 0.06}$	0.36 ± 0.06
Frequency features	0.30 ± 0.05	0.37 ± 0.06	0.36 ± 0.06	0.36 ± 0.06
CA indices	$\underline{0.29 \pm 0.05}$	$\mathbf{0.35 \pm 0.05}$	0.39 ± 0.34	0.36 ± 0.06
Ext. CA indices	$\mathbf{0.29 \pm 0.05}$	0.36 ± 0.06	$\mathbf{0.36 \pm 0.06}$	$\mathbf{0.36 \pm 0.06}$
Huser et al.	4185.74 ± 14696.02	0.40 ± 0.06	0.41 ± 0.07	0.45 ± 0.08
ICP pulse morphology	0.30 ± 0.05	0.37 ± 0.07	0.36 ± 0.06	$\underline{0.36 \pm 0.06}$
History of target	0.30 ± 0.05	0.36 ± 0.06	0.36 ± 0.06	0.36 ± 0.07
SAX BoW	0.33 ± 0.06	0.38 ± 0.06	0.42 ± 0.13	0.40 ± 0.07
ICP-ABP trace	0.31 ± 0.05	0.37 ± 0.06	0.39 ± 0.07	0.40 ± 0.07

Figure 5.4: Average root mean squared error (RMSE) and standard deviation of all models and the baseline model when predicting Pressure Reactivity Index (PRx).

(a) Mean absolute error.

(b) Root mean squared error.
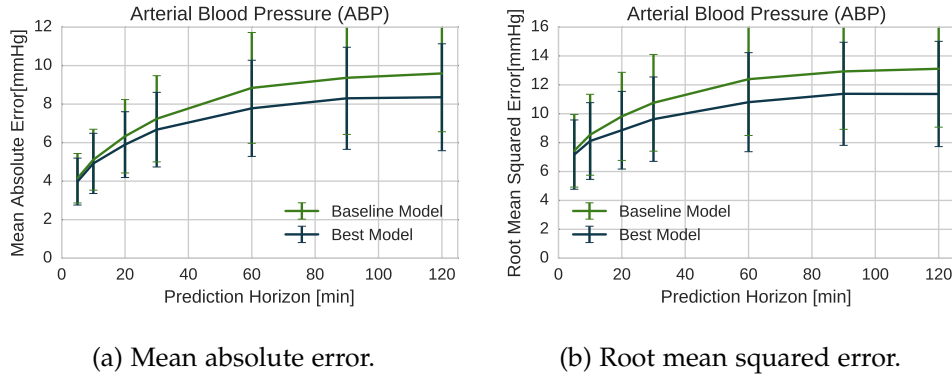
Figure 5.5: Forecasting Arterial Blood Pressure (ABP), the two plots show the prediction accuracy over time comparing the baseline model with the best model at each time point. The error bars indicate one standard deviation. The two tables below list the feature set of the best model and the precise error numbers, again with standard deviation.

(a) Mean absolute error.

(b) Root mean squared error.
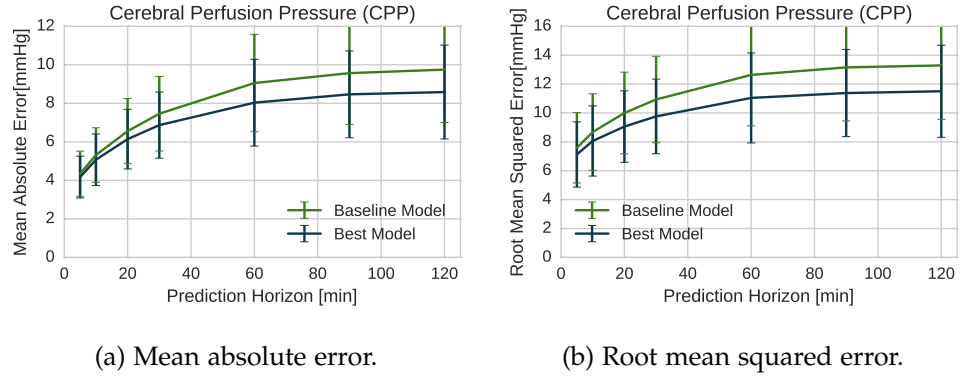
Figure 5.6: Forecasting Cerebral Perfusion Pressure (CPP), the two plots show the prediction accuracy over time comparing the baseline model with the best model at each time point. The error bars indicate one standard deviation. The two tables below list the feature set of the best model and the precise error numbers, again with standard deviation.

(a) Mean absolute error.

(b) Root mean squared error.
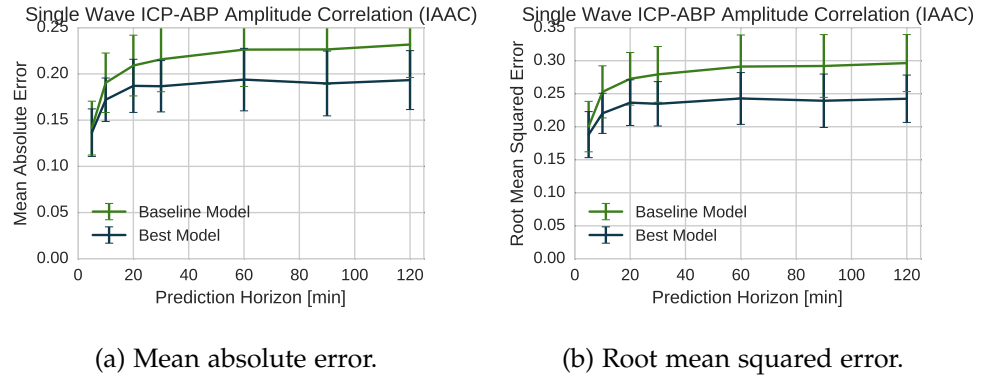
Figure 5.7: Forecasting Single Wave ICP-ABP Amplitude Correlation (IAAC), the two plots show the prediction accuracy over time comparing the baseline model with the best model at each time point. The error bars indicate one standard deviation. The two tables below list the feature set of the best model and the precise error numbers, again with standard deviation.

(a) Mean absolute error.

(b) Root mean squared error.
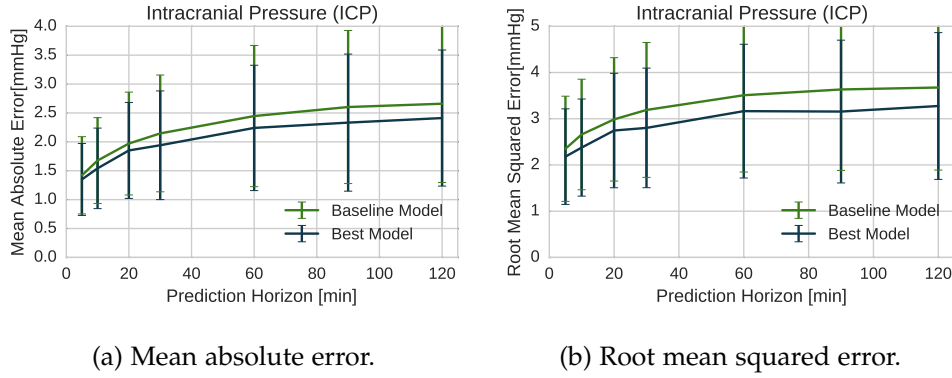
Figure 5.8: Forecasting Intracranial Pressure (ICP), the two plots show the prediction accuracy over time comparing the baseline model with the best model at each time point. The error bars indicate one standard deviation. The two tables below list the feature set of the best model and the precise error numbers, again with standard deviation.

(a) Mean absolute error.

(b) Root mean squared error.
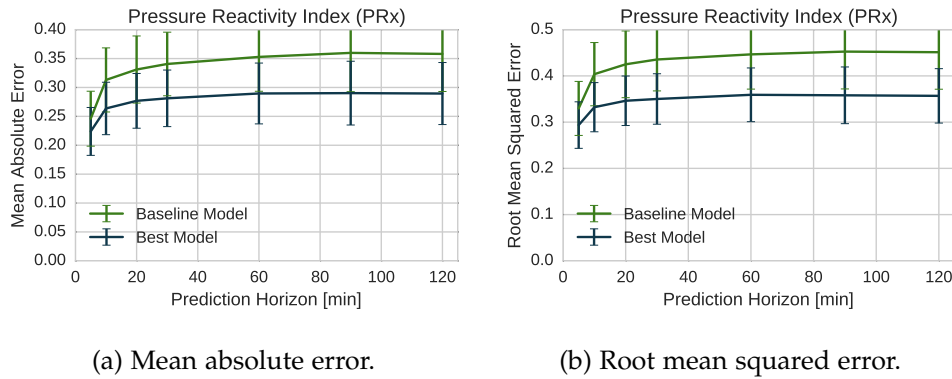
Figure 5.9: Forecasting Pressure Reactivity Index (PRx), the two plots show the prediction accuracy over time comparing the baseline model with the best model at each time point. The error bars indicate one standard deviation. The two tables below list the feature set of the best model and the precise error numbers, again with standard deviation.

(a) Mean absolute error.   (b) Root mean squared error.
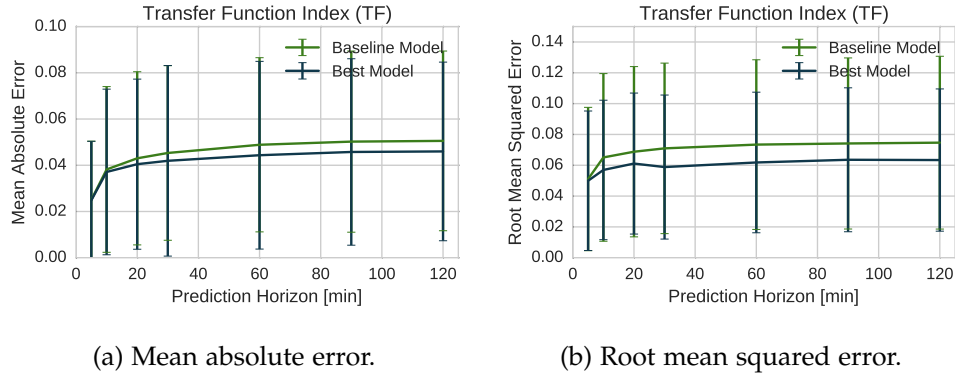
Figure 5.10: Forecasting Transfer Function Index (TF), the two plots show the prediction accuracy over time comparing the baseline model with the best model at each time point. The error bars indicate one standard deviation. The two tables below list the feature set of the best model and the precise error numbers, again with standard deviation.

5.3   Cambridge

We followed the same principles when evaluating our models on the Cambridge data set. We also did a leave-one-patient-out cross-validation with all records from the Cambridge data set containing ECG recordings. We also computed the MAE and the RMSE in this cross-validation after we did a hyperparameter search. The best model was then selected in the same way as in for the MIMIC II data set by selecting the model with the lowest macro average on MAE or RMSE[4].

We numbered the record from 0 up to 10. Since only 5 records had ECG information the resulting setup was:

$$\text{Horizons } T = \{5, 10, 20, 30, 60, 90, 120\}$$
$$\text{Feature Sets } F = \{history, stats, frequency, trace, index, sax, huser, morph\}$$
$$\text{Targets } P = \{ICP, ABP, CPP, PRx, TF\}$$
$$\text{Records } R = \{00, 01, 02, 03, 08\}$$

We used the same feature sets as described in Section 5.1.1. However, we removed all features the contained the signals HR, RESP, and SpO2 because they were not available. We also removed many more features from the

---

[4]Again, please keep in mind that the models were trained with the epsilon-insensitive absolute loss. Thus a higher accuracy in terms of RMSE is most likely achievable if the models would be trained with a squared loss.

Feature Set	5 min	30 min	60 min	120 min
Baseline	1.51 ± 1.34	3.47 ± 3.48	4.36 ± 4.13	3.55 ± 2.58
All features	1.53 ± 1.33	$\underline{3.45 \pm 3.43}$	4.39 ± 4.15	20.24 ± 37.45
Statistical summaries	1.51 ± 1.34	3.47 ± 3.46	4.27 ± 3.44	3.53 ± 2.66
Ext. stat. summaries	1.51 ± 1.34	3.47 ± 3.46	4.36 ± 3.38	3.59 ± 2.60
Frequency features	$\underline{1.51 \pm 1.34}$	$\mathbf{3.42 \pm 3.31}$	4.18 ± 3.40	$\mathbf{3.44 \pm 2.23}$
CA indices	1.51 ± 1.34	3.47 ± 3.47	$\underline{4.11 \pm 3.72}$	$\underline{3.48 \pm 2.25}$
Ext. CA indices	1.51 ± 1.34	3.47 ± 3.46	4.24 ± 4.04	3.55 ± 2.43
Huser et al.	1.51 ± 1.34	3.47 ± 3.46	4.37 ± 4.14	3.49 ± 2.62
ICP pulse morphology	1.51 ± 1.34	3.59 ± 3.36	$\mathbf{4.05 \pm 3.64}$	3.55 ± 2.27
History of target	1.52 ± 1.35	3.87 ± 3.62	4.43 ± 3.76	3.50 ± 2.66
SAX BoW	1.52 ± 1.34	3.47 ± 3.49	4.36 ± 4.15	3.52 ± 2.58
ICP-ABP trace	$\mathbf{1.51 \pm 1.34}$	3.47 ± 3.46	5.09 ± 3.29	3.50 ± 2.62

Figure 5.11: Average mean absolute error (MAE) and standard deviation of all models and the baseline model when predicting Intracranial Pressure (ICP).

feature set of Hüser et al. [28] since they include many other signals that are not available in the Cambridge data set.

Prediction Accuracy Over Time

Similar to the experiments for the MIMIC II data set we have summarized the results of our experiments for the Cambridge data set in Figure 5.15 for ABP, in Figure 5.16 for CPP, in Figure 5.18 for ICP, in Figure 5.17 for IAAC, in Figure 5.19 for PRx, and in Figure 5.20 for TF. The summary tables of the prediction accuracy of the different models can be found in Table 5.11 and 5.12 for ICP and in Table 5.13 and 5.14 for PRx. The remaining accuracy tables can be found in the appendix.

Feature Set	5 min	30 min	60 min	120 min
Baseline	2.46 ± 2.05	4.67 ± 4.24	5.66 ± 4.88	4.54 ± 3.08
All features	2.47 ± 2.05	4.65 ± 4.18	5.68 ± 4.91	30.47 ± 58.36
Statistical summaries	2.46 ± 2.05	4.66 ± 4.23	5.63 ± 3.83	4.52 ± 3.14
Ext. stat. summaries	2.46 ± 2.05	4.66 ± 4.23	5.70 ± 3.93	4.61 ± 3.03
Frequency features	2.46 ± 2.05	**4.60 ± 3.97**	5.51 ± 3.96	4.54 ± 2.76
CA indices	2.46 ± 2.05	4.66 ± 4.23	5.31 ± 4.28	4.55 ± 2.76
Ext. CA indices	2.46 ± 2.05	4.66 ± 4.23	5.48 ± 4.72	4.59 ± 2.90
Huser et al.	**2.46 ± 2.05**	4.66 ± 4.23	5.67 ± 4.87	**4.47 ± 3.09**
ICP pulse morphology	2.46 ± 2.05	4.78 ± 3.99	**5.22 ± 4.22**	4.64 ± 2.74
History of target	2.47 ± 2.06	5.11 ± 4.38	5.69 ± 4.39	4.53 ± 3.14
SAX BoW	2.46 ± 2.05	4.68 ± 4.25	5.63 ± 4.89	4.56 ± 3.01
ICP-ABP trace	2.46 ± 2.05	4.66 ± 4.23	6.35 ± 3.67	4.48 ± 3.10

Figure 5.12: Average root mean squared error (RMSE) and standard deviation of all models and the baseline model when predicting Intracranial Pressure (ICP).

Feature Set	5 min	30 min	60 min	120 min
Baseline	0.27 ± 0.04	0.39 ± 0.02	0.39 ± 0.06	0.42 ± 0.08
All features	0.27 ± 0.04	0.39 ± 0.02	0.40 ± 0.06	0.42 ± 0.08
Statistical summaries	0.27 ± 0.04	0.38 ± 0.01	0.36 ± 0.05	0.37 ± 0.07
Ext. stat. summaries	0.27 ± 0.04	0.40 ± 0.02	0.39 ± 0.05	0.43 ± 0.06
Frequency features	0.28 ± 0.07	0.42 ± 0.05	0.43 ± 0.08	0.43 ± 0.05
CA indices	0.26 ± 0.03	0.37 ± 0.02	0.37 ± 0.07	0.38 ± 0.06
Ext. CA indices	0.26 ± 0.03	0.41 ± 0.07	0.40 ± 0.04	0.42 ± 0.04
Huser et al.	0.27 ± 0.04	0.42 ± 0.05	0.39 ± 0.06	0.42 ± 0.08
ICP pulse morphology	0.33 ± 0.18	0.38 ± 0.02	0.36 ± 0.05	0.38 ± 0.07
History of target	**0.25 ± 0.03**	**0.33 ± 0.04**	**0.34 ± 0.03**	**0.36 ± 0.04**
SAX BoW	0.26 ± 0.04	0.39 ± 0.02	0.39 ± 0.06	0.42 ± 0.08
ICP-ABP trace	0.27 ± 0.04	0.39 ± 0.02	0.39 ± 0.06	0.42 ± 0.08

Figure 5.13: Average mean absolute error (MAE) and standard deviation of all models and the baseline model when predicting Pressure Reactivity Index (PRx).

Feature Set	5 min	30 min	60 min	120 min
Baseline	0.36 ± 0.05	0.50 ± 0.02	0.52 ± 0.06	0.53 ± 0.08
All features	0.36 ± 0.05	0.50 ± 0.02	0.52 ± 0.06	0.53 ± 0.08
Statistical summaries	0.36 ± 0.05	0.49 ± 0.04	0.46 ± 0.06	$\underline{0.46 \pm 0.08}$
Ext. stat. summaries	0.36 ± 0.05	0.50 ± 0.02	0.50 ± 0.07	0.54 ± 0.08
Frequency features	0.41 ± 0.15	0.66 ± 0.21	0.73 ± 0.47	0.62 ± 0.13
CA indices	$\underline{0.35 \pm 0.05}$	$\underline{0.47 \pm 0.02}$	$\underline{0.46 \pm 0.08}$	0.49 ± 0.06
Ext. CA indices	0.36 ± 0.04	0.52 ± 0.09	0.49 ± 0.04	0.53 ± 0.04
Huser et al.	0.36 ± 0.05	0.55 ± 0.10	0.52 ± 0.06	0.53 ± 0.08
ICP pulse morphology	0.49 ± 0.34	0.49 ± 0.01	0.46 ± 0.06	0.48 ± 0.09
History of target	$\mathbf{0.33 \pm 0.04}$	$\mathbf{0.41 \pm 0.04}$	$\mathbf{0.42 \pm 0.04}$	$\mathbf{0.44 \pm 0.05}$
SAX BoW	0.36 ± 0.05	0.50 ± 0.02	0.52 ± 0.06	0.53 ± 0.08
ICP-ABP trace	0.36 ± 0.05	0.50 ± 0.02	0.52 ± 0.06	0.53 ± 0.08

Figure 5.14: Average root mean squared error (RMSE) and standard deviation of all models and the baseline model when predicting Pressure Reactivity Index (PRx).

(a) Mean absolute error.

(b) Root mean squared error.
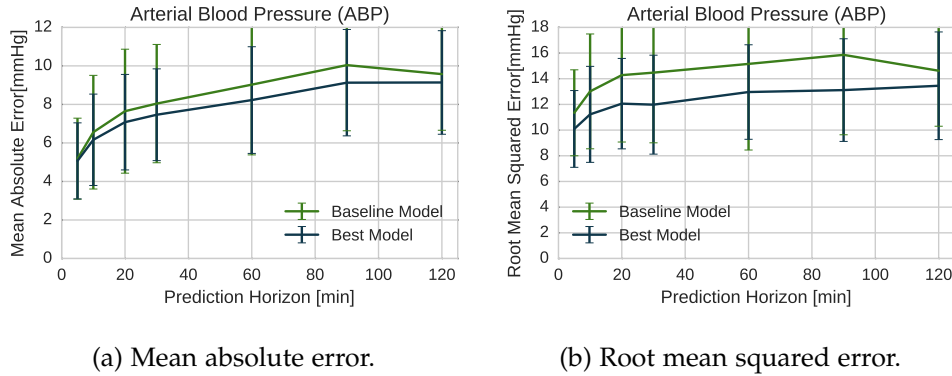
Figure 5.15: Forecasting Arterial Blood Pressure (ABP), the two plots show the prediction accuracy over time comparing the baseline model with the best model at each time point. The error bars indicate one standard deviation. The two tables below list the feature set of the best model and the precise error numbers, again with standard deviation.

(a) Mean absolute error.

(b) Root mean squared error.
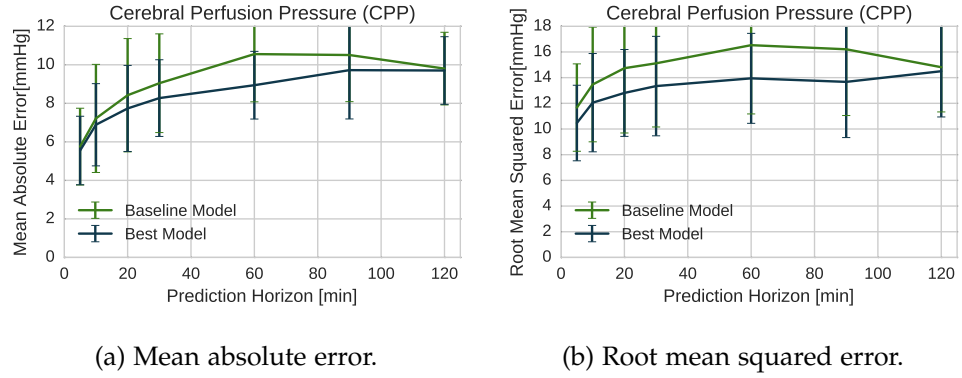
Figure 5.16: Forecasting Cerebral Perfusion Pressure (CPP), the two plots show the prediction accuracy over time comparing the baseline model with the best model at each time point. The error bars indicate one standard deviation. The two tables below list the feature set of the best model and the precise error numbers, again with standard deviation.

(a) Mean absolute error.

(b) Root mean squared error.
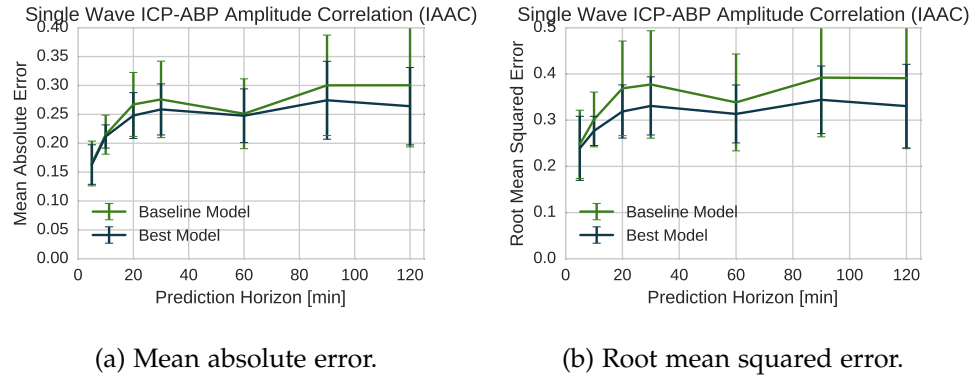
Figure 5.17: Forecasting Single Wave ICP-ABP Amplitude Correlation (IAAC), the two plots show the prediction accuracy over time comparing the baseline model with the best model at each time point. The error bars indicate one standard deviation. The two tables below list the feature set of the best model and the precise error numbers, again with standard deviation.

(a) Mean absolute error.

(b) Root mean squared error.
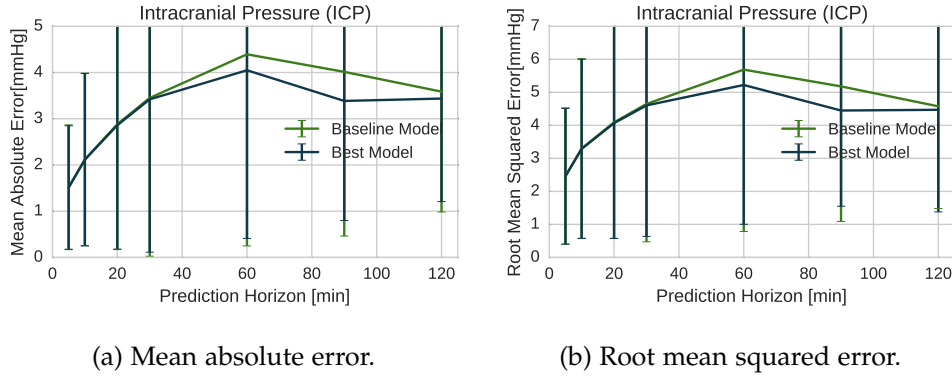
Figure 5.18: Forecasting Intracranial Pressure (ICP), the two plots show the prediction accuracy over time comparing the baseline model with the best model at each time point. The error bars indicate one standard deviation. The two tables below list the feature set of the best model and the precise error numbers, again with standard deviation.

(a) Mean absolute error.

(b) Root mean squared error.
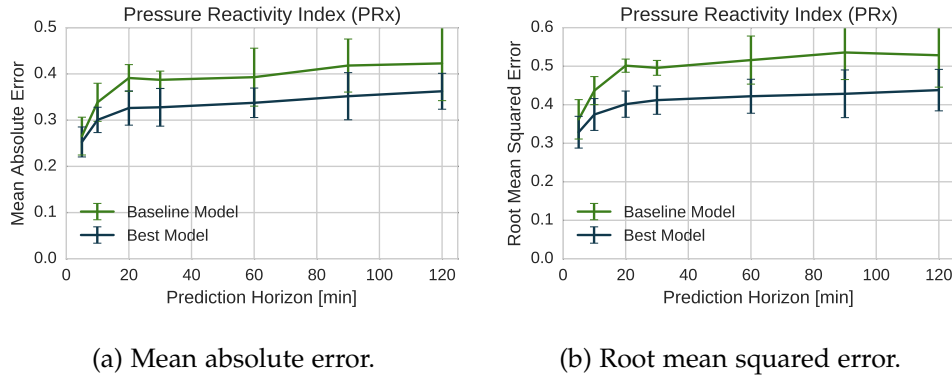
Figure 5.19: Forecasting Pressure Reactivity Index (PRx), the two plots show the prediction accuracy over time comparing the baseline model with the best model at each time point. The error bars indicate one standard deviation. The two tables below list the feature set of the best model and the precise error numbers, again with standard deviation.

(a) Mean absolute error.

(b) Root mean squared error.
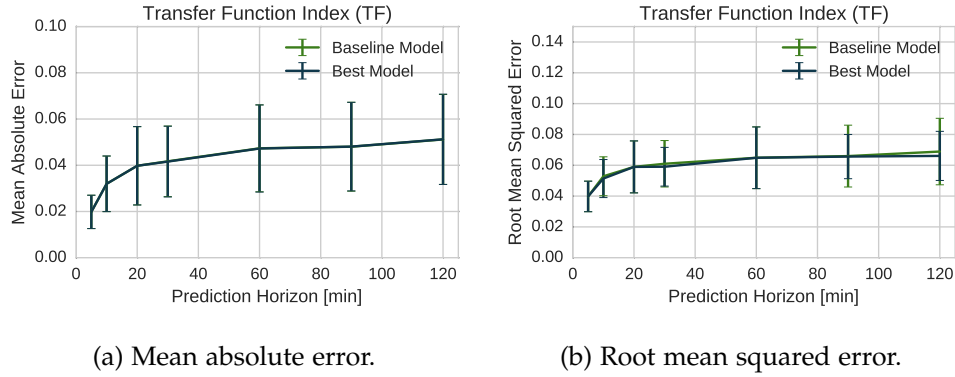
Figure 5.20: Forecasting Transfer Function Index (TF), the two plots show the prediction accuracy over time comparing the baseline model with the best model at each time point. The error bars indicate one standard deviation. The two tables below list the feature set of the best model and the precise error numbers, again with standard deviation.

5.4 Discussion

5.4.1 MIMIC II

We will first discuss the results achieved on the physiological targets and then the results achieved on the CA index targets.

When forecasting the physiological targets (ABP, CPP, and ICP) the macro average MAE and RMSE of the best model is lower than the baseline error. However, in all cases the baseline is well within one standard deviation. Thus, it is difficult to argue that our model performs significantly better than the baseline model. In most cases the feature set proposed by Hüser et al. [28] and the 5-minute history of the target performs best for both the MAE and the RMSE. We assume that they contain more information on the raw input signals and are thus better at forecasting the raw input signals. Most likely, the other feature sets either contain to little information (statistical summaries, frequency) or too much non-relevant information (CA indices, SAX encoding, morphology).

When forecasting IAAC and PRx our achieved macro average MAE and RMSE of the best model is more than one standard deviation lower than the baseline error. It is thus easier but still difficult to argue that our model is better than the baseline. When predicting IAAC, the statistical summaries performed the best. When predicting PRx the extended CA indices performed best. We are surprised that the extended CA indices are not best for predicting both targets and think that this fact also needs further investigation.

When forecasting the TF index we could not achieve a significant increase in accuracy compared to the baseline model. We assume that this is partly because of the way we compute the TF index. Since we were not able to validate our computed TF index the same way other CA indices have been validate (by correlation with outcome) we conclude that further investigation into computation of the TF index is necessary.

For all prediction targets we see that the best feature set does depend on the prediction target. Physiological targets are better predicted by the features proposed in [28] or the raw history, CA indices are best predicted by statistical summaries and CA indices. We think that the bad performance of the morphological features needs further investigation and we also think that is necessary to analyze the feature sets more in detail to remove features that are not contributing to prediction accuracy. Currently we assume that most feature sets are just too noisy and that a combination of the most predictive features of all individual feature sets we achieve a significantly higher accuracy.

5.4.2 Cambridge

The Cambridge data set evaluated almost the same way as the MIMIC II data set. We also observe the low prediction accuracy for the targets ABP, CPP, ICP, and TF. We also see an increase in prediction accuracy in PRx and IAAC. Finally, we see similar feature sets making the most accurate prediction for the different prediction targets and the different prediction horizons. One significant difference is, that the features proposed in [28] do not achieve the highest accuracy anymore. This is most likely because many features in this set could not be computed due to missing signals.

5.4.3 Comparison to Hüser et al.

We included the feature set of Hüser et al. [28] as one of our evaluated features sets. It showed to be an important feature set for predicting the physiological parameters ABP, CPP, and ICP. For all three of them the features proposed by Hüser et al. achieved the lowest MAE for the clinically relevant forecasting horizons of 30 minutes or more. However, when predicting CA indices the feature set containing statistical summaries or (extended) CA indices achieved a higher accuracy.

5.4.4 Comparison to Kashif et al.

Kashif et al. [30] have reported a bias of $1.5mmHg \pm 5.9mmHg$ for their non-invasive ICP estimation model based on Transcranial Doppler readings of the CBFV. The bias is computed as the average difference between the true and the predicted value.

For our models we only computed the MAE which will be bigger or equal to the absolute bias. Our best model on a prediction horizon of 5 minutes achieved a MAE of $1.35mmHg \pm 0.62mmHg$. However, already our baseline model achieved a MAE of $1.42mmHg \pm 0.67mmHg$. This shows that non-invasively estimating ICP is a lot harder than predicting it 5 minutes into the future.

5.4.5 Comparison to Zhang at al.

Zhang et al. [56] have reported accuracy results for their model which predicts the mean ICP from the current time up to a time of 45 minutes. They achieved an R^2 score of 0.93 ± 0.05 (0.81 ± 0.11, 0.56 ± 0.25) for the time horizon $T = 15min$ ($30min$, $45min$). They also report an MSE of $0.88mmHg \pm 0.58mmHg$ ($3.26mmHg \pm 1.96mmHg$, $8.12mmHg \pm 4.72mmHg$) and an RAE of $9\% \pm 3\%$ ($24\% \pm 11\%$, $49\% \pm 23\%$) respectively.

In contrast to them we predicted the 30-second mean ICP at the forecasting horizon T. This is a much harder problem since the variance of our prediction target increases the longer the prediction horizon is and the variance of their prediction target decreases. Still, we compare their results to our achieved performance at $T = 10min$ and $T = 30min$. We achieved an R^2 score of 0.19 ± 0.05 (0.22 ± 0.07), an MSE of $6.71mmHg \pm 7.75mmHg$ ($9.46mmHg \pm 10.96mmHg$), and an RAE of $92\% \pm 4\%$ ($91\% \pm 6\%$) in our leave-one-patient-out cross-validation.

Chapter 6

Conclusion

We proposed different models to predict the physiological parameters ABP, CPP, and ICP and the cerebral autoregulation (CA) indices PRx, TF, and IAAC. We evaluated the different models on prediction horizons up to 2 hours and found that the set of best features depend on the prediction target and sometimes also on the prediction horizon. When predicting physiological targets, the best performing feature set was proposed in [28] followed by the 5-minute history of the target value. We achieved a relative decrease of prediction error by up to 11% (13%, 13%) for ICP (ABP, CPP). When predicting CA indices, the best performing feature set were the statistical summaries and the CA indices. We achieved a relative decrease of prediction error by up to 24% and 21%, for PRx and IAAC respectively.

Large scale prediction of all patients clinical state inside an intensive care unit could increase effectiveness and efficiency of treatment. Some neuro-intensive-care units already have about 100 beds. This requires the doctors to work at larger scales and it increases the risk of doctors overlooking critical information important to treat a single patient. By providing the doctors with an early warning system that independently monitors the patients and alerts the doctors when a vital parameter is expected to worsen, the doctors can more efficiently treat all patients.

We have proposed predictive models for such a system. However, we also showed that a lot of work still needs to be done to increase the prediction accuracy of those models. We hope that this work has provided a starting point for a more detailed analysis and extended research on forecasting not only basic physiological parameters but also clinical indices to move from reactive treatment methods to more proactive treatment methods.

6.1 Future Work

During this work we encountered many problems and questions that we could not investigate but which appeared to be interesting research topics:

- The concept of defining a patient specific threshold struck us to be very important. However, the current method is lacking due to the flat or concave shape of the CPP-Index curve. To make those thresholds relevant in an intensive care unit one would need to find a more robust way of finding the threshold values. Maybe it is possible to fit a general model to the patient specific context based on very little physiological information. This model could then increasingly refined based on monitoring results.

- We only had access to 26 patients in our biggest data set. This was due to the fact that the MIMIC II data set contains only few recordings of patients suffering a neurological condition. When working with models which have many complex features this is often too little to both train and evaluate the model. We would be very interested in doing an evaluation of our proposed models on a much bigger data set.

- In our work we did a coarse grained analysis on the predictive power of feature sets. However, it would be interesting to also evaluate the predictive power of the individual features. For this it would be necessary analyze the weights put by the linear model to find a starting set of the most promising features. From that on one could iteratively remove one or many features until the model accuracy does not improve anymore. One could also take the most predictive features of each set and collect them into one set.

- It would also be interesting to investigate why especially the physiological targets have such a high variance even when averaged over 30 seconds. Maybe models from financial markets are applicable as well since the targets often seem to be zero mean.

- We proposed an algorithm to convert the three properties of the transfer function into one value. We think that this algorithm needs improvement because we were not able to predict it accurately and we think that the algorithm also needs validation in a similar way to how other CA indices have been validated. Also, one would need to investigate if it does not make more sense to predict all three parameters individually instead of combining them into one value.

- The MIMIC II database also contains clinical information on a matched subset of database records. We propose to further investigate the use of such clinical information. However, we were not able to get enough

matched records to incorporate clinical information into our predictive models. Thus, using this clinical information is not possible for the MIMIC II data set except if additional records containing ICP recordings are matched with clinical records.

Appendix A

Appendix

A.1 MIMIC II

Feature Set	5 min	30 min	60 min	120 min
Baseline	4.15 ± 1.28	7.24 ± 2.24	8.84 ± 2.88	9.59 ± 3.03
Statistical summaries	4.12 ± 1.27	7.25 ± 2.24	8.85 ± 2.88	9.34 ± 2.90
Ext. stat. summaries	4.10 ± 1.25	$\underline{6.87 \pm 2.12}$	8.47 ± 2.72	$\underline{8.80 \pm 2.89}$
Frequency features	4.12 ± 1.28	7.09 ± 2.21	8.54 ± 2.83	9.14 ± 2.85
CA indices	4.14 ± 1.28	7.22 ± 2.22	8.84 ± 2.86	9.42 ± 2.94
Ext. CA indices	$\mathbf{3.98 \pm 1.22}$	7.00 ± 2.40	8.51 ± 2.75	9.10 ± 2.82
Huser et al.	$\underline{4.08 \pm 1.26}$	$\mathbf{6.67 \pm 1.94}$	$\mathbf{7.78 \pm 2.50}$	$\mathbf{8.36 \pm 2.78}$
ICP pulse morphology	4.13 ± 1.28	7.20 ± 2.21	8.83 ± 2.87	9.19 ± 2.91
History of target	4.09 ± 1.23	7.00 ± 2.12	$\underline{8.37 \pm 2.88}$	9.07 ± 2.82
SAX BoW	4.15 ± 1.28	7.24 ± 2.22	8.84 ± 2.88	9.59 ± 3.03
ICP-ABP trace	4.15 ± 1.28	7.16 ± 2.22	8.84 ± 2.88	9.42 ± 3.01

Figure A.1: Average mean absolute error (MAE) and standard deviation of all models and the baseline model when predicting Arterial Blood Pressure (ABP).

Feature Set	5 min	30 min	60 min	120 min
Baseline	7.44 ± 2.51	10.76 ± 3.35	12.39 ± 3.90	13.11 ± 4.04
Statistical summaries	7.34 ± 2.49	10.77 ± 3.35	12.41 ± 3.89	12.65 ± 3.84
Ext. stat. summaries	7.31 ± 2.47	$\underline{9.92 \pm 3.08}$	11.75 ± 3.66	$\underline{11.67 \pm 3.63}$
Frequency features	7.39 ± 2.51	10.47 ± 3.28	11.84 ± 3.76	12.27 ± 3.67
CA indices	7.42 ± 2.51	10.74 ± 3.33	12.39 ± 3.89	12.72 ± 3.85
Ext. CA indices	11.11 ± 21.84	30.01 ± 103.77	11.82 ± 3.69	12.36 ± 3.76
Huser et al.	$\mathbf{7.17 \pm 2.40}$	$\mathbf{9.62 \pm 2.92}$	$\mathbf{10.80 \pm 3.43}$	$\mathbf{11.37 \pm 3.64}$
ICP pulse morphology	7.39 ± 2.51	10.66 ± 3.30	12.38 ± 3.90	12.37 ± 3.78
History of target	$\underline{7.22 \pm 2.40}$	10.11 ± 3.05	$\underline{11.43 \pm 3.69}$	12.24 ± 3.65
SAX BoW	7.44 ± 2.51	10.76 ± 3.33	12.40 ± 3.90	13.12 ± 4.05
ICP-ABP trace	7.44 ± 2.51	10.64 ± 3.32	12.40 ± 3.90	12.80 ± 4.00

Figure A.2: Average root mean squared error (RMSE) and standard deviation of all models and the baseline model when predicting Arterial Blood Pressure (ABP).

Feature Set	5 min	30 min	60 min	120 min
Baseline	4.35 ± 1.17	7.46 ± 1.94	9.05 ± 2.53	9.75 ± 2.75
Statistical summaries	4.34 ± 1.17	7.36 ± 1.90	8.88 ± 2.58	9.68 ± 2.71
Ext. stat. summaries	$\underline{4.18 \pm 1.09}$	7.05 ± 1.86	8.36 ± 2.50	8.91 ± 2.66
Frequency features	4.33 ± 1.16	7.47 ± 1.93	8.80 ± 2.47	9.37 ± 2.60
CA indices	4.35 ± 1.17	7.43 ± 1.94	9.04 ± 2.51	9.50 ± 2.70
Ext. CA indices	$\mathbf{4.17 \pm 1.08}$	7.11 ± 2.00	8.43 ± 2.50	9.03 ± 2.70
Huser et al.	4.19 ± 1.11	$\mathbf{6.87 \pm 1.72}$	$\mathbf{8.03 \pm 2.25}$	$\mathbf{8.58 \pm 2.44}$
ICP pulse morphology	4.35 ± 1.17	7.44 ± 1.93	8.79 ± 2.51	9.29 ± 2.69
History of target	4.25 ± 1.13	7.11 ± 1.92	8.59 ± 2.50	9.24 ± 2.71
SAX BoW	4.35 ± 1.17	10.10 ± 12.45	9.04 ± 2.53	9.74 ± 2.74
ICP-ABP trace	4.35 ± 1.17	7.37 ± 1.93	8.88 ± 2.52	9.57 ± 2.81

Figure A.3: Average mean absolute error (MAE) and standard deviation of all models and the baseline model when predicting Cerebral Perfusion Pressure (CPP).

Feature Set	5 min	30 min	60 min	120 min
Baseline	7.59 ± 2.43	10.93 ± 2.99	12.63 ± 3.53	13.29 ± 3.73
Statistical summaries	7.58 ± 2.43	10.60 ± 2.88	12.32 ± 3.77	13.12 ± 3.67
Ext. stat. summaries	$\mathbf{7.13 \pm 2.26}$	10.94 ± 5.75	$\underline{11.33 \pm 3.28}$	$\underline{11.73 \pm 3.29}$
Frequency features	7.46 ± 2.39	10.94 ± 2.99	12.14 ± 3.35	12.60 ± 3.42
CA indices	7.58 ± 2.43	10.84 ± 2.96	12.56 ± 3.50	12.78 ± 3.54
Ext. CA indices	11.25 ± 21.88	21.40 ± 58.87	11.43 ± 3.29	11.85 ± 3.32
Huser et al.	$\underline{7.13 \pm 2.26}$	$\mathbf{9.76 \pm 2.59}$	$\mathbf{11.04 \pm 3.12}$	$\mathbf{11.50 \pm 3.19}$
ICP pulse morphology	7.59 ± 2.43	10.87 ± 2.96	12.18 ± 3.51	12.45 ± 3.48
History of target	7.30 ± 2.33	$\underline{10.20 \pm 2.80}$	11.83 ± 3.32	12.37 ± 3.44
SAX BoW	7.59 ± 2.43	14.32 ± 16.07	12.63 ± 3.54	13.29 ± 3.72
ICP-ABP trace	7.59 ± 2.43	10.79 ± 2.99	12.38 ± 3.51	13.02 ± 3.72

Figure A.4: Average root mean squared error (RMSE) and standard deviation of all models and the baseline model when predicting Cerebral Perfusion Pressure (CPP).

Feature Set	5 min	30 min	60 min	120 min
Baseline	0.03 ± 0.03	0.05 ± 0.04	0.05 ± 0.04	0.05 ± 0.04
Statistical summaries	0.02 ± 0.03	$\underline{0.04 \pm 0.04}$	0.04 ± 0.04	$\mathbf{0.05 \pm 0.04}$
Ext. stat. summaries	0.02 ± 0.03	$\mathbf{0.04 \pm 0.04}$	0.05 ± 0.04	0.05 ± 0.04
Frequency features	$\underline{0.02 \pm 0.03}$	0.04 ± 0.04	$\underline{0.04 \pm 0.04}$	0.05 ± 0.04
CA indices	0.03 ± 0.03	0.04 ± 0.04	0.05 ± 0.04	0.05 ± 0.04
Ext. CA indices	0.03 ± 0.03	0.29 ± 1.28	$\mathbf{0.04 \pm 0.04}$	750.78 ± 3827.13
Huser et al.	0.03 ± 0.04	0.05 ± 0.04	0.05 ± 0.04	0.05 ± 0.04
ICP pulse morphology	0.02 ± 0.03	0.04 ± 0.04	0.05 ± 0.04	$\underline{0.05 \pm 0.04}$
History of target	$\mathbf{0.02 \pm 0.03}$	0.04 ± 0.04	0.05 ± 0.04	0.05 ± 0.04
SAX BoW	0.03 ± 0.03	0.18 ± 0.70	0.05 ± 0.04	0.05 ± 0.04
ICP-ABP trace	0.03 ± 0.03	0.04 ± 0.04	0.05 ± 0.04	0.10 ± 0.16

Figure A.5: Average mean absolute error (MAE) and standard deviation of all models and the baseline model when predicting Transfer Function Index (TF).

Feature Set	5 min	30 min	60 min	120 min
Baseline	0.05 ± 0.05	0.07 ± 0.06	0.07 ± 0.06	0.07 ± 0.06
Statistical summaries	$\underline{0.05 \pm 0.05}$	0.06 ± 0.05	0.06 ± 0.05	$\underline{0.06 \pm 0.05}$
Ext. stat. summaries	0.05 ± 0.05	$\mathbf{0.06 \pm 0.05}$	0.07 ± 0.05	0.07 ± 0.05
Frequency features	0.05 ± 0.05	0.06 ± 0.05	$\underline{0.06 \pm 0.04}$	0.07 ± 0.05
CA indices	0.05 ± 0.05	0.06 ± 0.05	0.07 ± 0.05	0.07 ± 0.05
Ext. CA indices	0.05 ± 0.05	2.28 ± 11.29	$\mathbf{0.06 \pm 0.05}$	1356.31 ± 6914.66
Huser et al.	0.06 ± 0.05	0.07 ± 0.06	0.07 ± 0.06	0.07 ± 0.06
ICP pulse morphology	0.05 ± 0.05	0.07 ± 0.05	0.06 ± 0.04	$\mathbf{0.06 \pm 0.05}$
History of target	$\mathbf{0.05 \pm 0.05}$	0.07 ± 0.05	0.07 ± 0.05	0.06 ± 0.05
SAX BoW	0.05 ± 0.05	0.25 ± 0.90	0.07 ± 0.06	0.07 ± 0.05
ICP-ABP trace	0.05 ± 0.05	$\underline{0.06 \pm 0.05}$	0.07 ± 0.05	0.12 ± 0.16

Figure A.6: Average root mean squared error (RMSE) and standard deviation of all models and the baseline model when predicting Transfer Function Index (TF).

Feature Set	5 min	30 min	60 min	120 min
Baseline	0.14 ± 0.03	0.22 ± 0.04	0.23 ± 0.04	0.23 ± 0.04
Statistical summaries	$\mathbf{0.14 \pm 0.03}$	$\underline{0.19 \pm 0.03}$	$\mathbf{0.19 \pm 0.03}$	$\mathbf{0.19 \pm 0.03}$
Ext. stat. summaries	0.14 ± 0.03	0.19 ± 0.03	0.20 ± 0.03	1.58 ± 6.96
Frequency features	0.14 ± 0.03	0.19 ± 0.03	0.20 ± 0.03	0.20 ± 0.03
CA indices	0.14 ± 0.03	0.20 ± 0.05	0.20 ± 0.03	0.20 ± 0.03
Ext. CA indices	$\underline{0.14 \pm 0.03}$	$\mathbf{0.19 \pm 0.03}$	0.20 ± 0.03	0.20 ± 0.03
Huser et al.	0.14 ± 0.03	0.22 ± 0.03	0.23 ± 0.04	0.23 ± 0.04
ICP pulse morphology	0.14 ± 0.03	0.19 ± 0.03	$\underline{0.19 \pm 0.03}$	$\underline{0.19 \pm 0.03}$
History of target	0.14 ± 0.03	0.19 ± 0.03	0.20 ± 0.04	0.20 ± 0.04
SAX BoW	0.14 ± 0.03	0.20 ± 0.03	0.22 ± 0.04	0.23 ± 0.04
ICP-ABP trace	0.15 ± 0.03	0.20 ± 0.03	0.21 ± 0.03	0.21 ± 0.03

Figure A.7: Average mean absolute error (MAE) and standard deviation of all models and the baseline model when predicting Single Wave ICP-ABP Amplitude Correlation (IAAC).

Feature Set	5 min	30 min	60 min	120 min
Baseline	0.20 ± 0.04	0.28 ± 0.04	0.29 ± 0.05	0.30 ± 0.04
Statistical summaries	$\mathbf{0.19 \pm 0.03}$	$\underline{0.23 \pm 0.04}$	$\mathbf{0.24 \pm 0.04}$	$\underline{0.24 \pm 0.04}$
Ext. stat. summaries	0.19 ± 0.04	0.24 ± 0.04	0.25 ± 0.04	8.59 ± 42.41
Frequency features	0.19 ± 0.04	0.24 ± 0.04	0.26 ± 0.04	0.25 ± 0.04
CA indices	0.20 ± 0.04	0.25 ± 0.06	0.25 ± 0.04	0.25 ± 0.03
Ext. CA indices	0.19 ± 0.04	$\mathbf{0.23 \pm 0.03}$	0.25 ± 0.04	0.25 ± 0.04
Huser et al.	0.20 ± 0.04	0.28 ± 0.04	0.29 ± 0.05	0.30 ± 0.04
ICP pulse morphology	0.19 ± 0.04	0.24 ± 0.04	$\underline{0.24 \pm 0.04}$	$\mathbf{0.24 \pm 0.04}$
History of target	$\underline{0.19 \pm 0.04}$	0.24 ± 0.04	0.25 ± 0.04	0.26 ± 0.05
SAX BoW	0.20 ± 0.04	0.26 ± 0.04	0.29 ± 0.05	0.29 ± 0.05
ICP-ABP trace	0.20 ± 0.04	0.25 ± 0.04	0.26 ± 0.04	0.27 ± 0.04

Figure A.8: Average root mean squared error (RMSE) and standard deviation of all models and the baseline model when predicting Single Wave ICP-ABP Amplitude Correlation (IAAC).

Record	Samples	ABP	CPP	IAAC	ICP	PRx	TF	Length
3106263	40900	54%	36%	35%	78%	36%	36%	4 days 17:36:40
3142868	50382	100%	96%	97%	96%	97%	97%	5 days 19:57:00
3148126	16146	100%	97%	97%	97%	98%	98%	1 days 20:51:00
3160820	6518	99%	93%	95%	93%	96%	96%	0 days 18:06:20
3169632	8230	99%	94%	97%	94%	97%	97%	0 days 22:51:40
3189000	10954	100%	96%	97%	96%	97%	97%	1 days 06:25:40
3270980	58753	99%	96%	97%	96%	98%	98%	6 days 19:12:10
3309132	18243	99%	97%	99%	97%	99%	99%	2 days 02:40:30
3319401	21752	99%	95%	96%	95%	97%	97%	2 days 12:25:20
3365681	49077	33%	33%	34%	99%	34%	34%	5 days 16:19:30
3367596	5804	95%	95%	95%	96%	95%	95%	0 days 16:07:20
3487247	4630	97%	89%	92%	91%	92%	92%	0 days 12:51:40
3516004	65777	98%	87%	89%	88%	89%	89%	7 days 14:42:50
3543187	24121	96%	96%	96%	96%	97%	97%	2 days 19:00:10
3562822	11282	100%	94%	98%	94%	98%	98%	1 days 07:20:20
3624651	8258	99%	92%	93%	93%	93%	93%	0 days 22:56:20
3629298	25412	99%	96%	98%	97%	98%	98%	2 days 22:35:20
3642023	10889	67%	66%	52%	99%	67%	67%	1 days 06:14:50
3655233	64167	98%	98%	93%	98%	99%	99%	7 days 10:14:30
3656395	5398	99%	82%	82%	82%	82%	82%	0 days 14:59:40
3668415	21682	87%	72%	77%	83%	77%	77%	2 days 12:13:40
3688532	34547	97%	94%	96%	96%	96%	96%	3 days 23:57:50
3693937	7018	99%	98%	99%	98%	99%	99%	0 days 19:29:40
3700665	7957	84%	80%	82%	96%	82%	82%	0 days 22:06:10
3774557	7844	99%	97%	98%	97%	98%	98%	0 days 21:47:20
3938777	7268	78%	78%	78%	99%	78%	78%	0 days 20:11:20

Table A.1: Availability of the individual targets at the *5 minutes* prediction horizon for each record from the MIMIC II data set.

Record	Samples	ABP	CPP	IAAC	ICP	PRx	TF	Length
3106263	40900	53%	35%	34%	78%	35%	35%	4 days 17:36:40
3142868	50382	99%	95%	96%	95%	96%	96%	5 days 19:57:00
3148126	16146	99%	95%	96%	95%	96%	96%	1 days 20:51:00
3160820	6518	97%	90%	92%	91%	93%	93%	0 days 18:06:20
3169632	8230	97%	91%	94%	91%	94%	94%	0 days 22:51:40
3189000	10954	98%	94%	94%	94%	95%	95%	1 days 06:25:40
3270980	58753	99%	95%	96%	95%	97%	97%	6 days 19:12:10
3309132	18243	98%	96%	97%	96%	97%	97%	2 days 02:40:30
3319401	21752	98%	93%	95%	94%	95%	95%	2 days 12:25:20
3365681	49077	33%	33%	33%	99%	33%	33%	5 days 16:19:30
3367596	5804	90%	90%	89%	90%	89%	89%	0 days 16:07:20
3487247	4630	93%	85%	88%	87%	88%	88%	0 days 12:51:40
3516004	65777	97%	86%	87%	87%	87%	87%	7 days 14:42:50
3543187	24121	95%	95%	95%	95%	95%	95%	2 days 19:00:10
3562822	11282	98%	93%	96%	93%	96%	96%	1 days 07:20:20
3624651	8258	97%	88%	88%	88%	89%	89%	0 days 22:56:20
3629298	25412	99%	96%	97%	96%	97%	97%	2 days 22:35:20
3642023	10889	65%	65%	50%	98%	66%	66%	1 days 06:14:50
3655233	64167	98%	97%	91%	98%	98%	98%	7 days 10:14:30
3656395	5398	96%	75%	74%	75%	74%	74%	0 days 14:59:40
3668415	21682	85%	69%	73%	81%	73%	73%	2 days 12:13:40
3688532	34547	96%	92%	94%	95%	94%	94%	3 days 23:57:50
3693937	7018	97%	96%	97%	96%	97%	97%	0 days 19:29:40
3700665	7957	80%	76%	78%	93%	78%	78%	0 days 22:06:10
3774557	7844	97%	95%	96%	96%	96%	96%	0 days 21:47:20
3938777	7268	76%	76%	76%	97%	76%	76%	0 days 20:11:20

Table A.2: Availability of the individual targets at the *30 minutes* prediction horizon for each record from the MIMIC II data set.

Record	Samples	ABP	CPP	IAAC	ICP	PRx	TF	Length
3106263	40900	51%	34%	33%	76%	34%	34%	4 days 17:36:40
3142868	50382	98%	94%	95%	94%	95%	95%	5 days 19:57:00
3148126	16146	95%	92%	92%	92%	92%	92%	1 days 20:51:00
3160820	6518	88%	82%	84%	82%	84%	84%	0 days 18:06:20
3169632	8230	91%	85%	88%	85%	88%	88%	0 days 22:51:40
3189000	10954	93%	89%	90%	90%	91%	91%	1 days 06:25:40
3270980	58753	98%	94%	95%	94%	96%	96%	6 days 19:12:10
3309132	18243	95%	93%	95%	93%	95%	95%	2 days 02:40:30
3319401	21752	96%	91%	93%	92%	93%	93%	2 days 12:25:20
3365681	49077	31%	31%	31%	98%	31%	31%	5 days 16:19:30
3367596	5804	81%	81%	80%	81%	80%	80%	0 days 16:07:20
3487247	4630	82%	75%	78%	77%	78%	78%	0 days 12:51:40
3516004	65777	96%	85%	86%	85%	86%	86%	7 days 14:42:50
3543187	24121	91%	90%	91%	91%	91%	91%	2 days 19:00:10
3562822	11282	94%	88%	91%	88%	91%	91%	1 days 07:20:20
3624651	8258	91%	82%	83%	82%	83%	83%	0 days 22:56:20
3629298	25412	96%	94%	95%	94%	95%	95%	2 days 22:35:20
3642023	10889	60%	60%	46%	93%	61%	61%	1 days 06:14:50
3655233	64167	97%	96%	89%	97%	97%	97%	7 days 10:14:30
3656395	5398	86%	63%	62%	63%	62%	62%	0 days 14:59:40
3668415	21682	81%	64%	69%	75%	69%	69%	2 days 12:13:40
3688532	34547	95%	90%	92%	93%	92%	92%	3 days 23:57:50
3693937	7018	89%	88%	89%	88%	89%	89%	0 days 19:29:40
3700665	7957	67%	63%	64%	86%	64%	64%	0 days 22:06:10
3774557	7844	90%	89%	89%	89%	89%	89%	0 days 21:47:20
3938777	7268	69%	68%	69%	89%	69%	69%	0 days 20:11:20

Table A.3: Availability of the individual targets at the *120 minutes* prediction horizon for each record from the MIMIC II data set.

Feature Set	5 min	30 min	60 min	120 min
Baseline	5.18 ± 2.06	8.00 ± 3.06	9.08 ± 3.52	9.70 ± 2.90
All features	5.09 ± 2.03	7.87 ± 2.93	9.03 ± 3.65	9.57 ± 2.91
Statistical summaries	5.19 ± 2.00	10.16 ± 6.93	8.97 ± 3.48	9.62 ± 3.04
Ext. stat. summaries	5.13 ± 1.78	7.97 ± 2.42	$\mathbf{8.23 \pm 2.77}$	$\mathbf{9.14 \pm 2.69}$
Frequency features	5.20 ± 1.99	8.26 ± 3.05	9.00 ± 2.55	9.62 ± 2.93
CA indices	5.19 ± 1.97	7.83 ± 3.14	8.98 ± 3.49	9.62 ± 2.98
Ext. CA indices	$\mathbf{5.07 \pm 1.97}$	20.69 ± 29.97	$\underline{8.34 \pm 2.82}$	$\underline{9.19 \pm 3.06}$
Huser et al.	$\underline{5.08 \pm 2.02}$	$\underline{7.61 \pm 2.91}$	9.09 ± 3.15	9.80 ± 2.52
ICP pulse morphology	5.14 ± 2.01	8.43 ± 3.34	8.97 ± 3.48	9.59 ± 3.05
History of target	5.29 ± 1.76	$\mathbf{7.46 \pm 2.38}$	8.85 ± 2.66	14.24 ± 5.52
SAX BoW	5.18 ± 1.97	7.90 ± 3.11	9.08 ± 3.52	9.70 ± 2.90
ICP-ABP trace	5.31 ± 1.91	8.32 ± 3.41	9.07 ± 3.52	9.70 ± 2.90

Table A.4: Average mean absolute error (MAE) and standard deviation of all models and the baseline model when predicting Arterial Blood Pressure (ABP).

Feature Set	5 min	30 min	60 min	120 min
Baseline	11.31 ± 3.33	14.41 ± 5.46	15.47 ± 6.29	14.79 ± 4.42
All features	10.96 ± 3.15	13.06 ± 4.75	15.15 ± 6.71	14.62 ± 4.32
Statistical summaries	10.95 ± 3.30	16.96 ± 11.03	14.90 ± 5.98	14.47 ± 4.79
Ext. stat. summaries	$\underline{10.31 \pm 3.26}$	$\underline{12.75 \pm 3.94}$	$\underline{13.38 \pm 4.70}$	$\mathbf{13.45 \pm 4.19}$
Frequency features	11.00 ± 3.31	13.62 ± 4.86	14.44 ± 4.75	14.48 ± 4.48
CA indices	10.96 ± 3.29	13.43 ± 5.14	15.05 ± 6.06	14.45 ± 4.65
Ext. CA indices	10.36 ± 3.42	35.87 ± 54.20	13.66 ± 4.91	$\underline{13.75 \pm 4.79}$
Huser et al.	10.39 ± 3.36	13.02 ± 5.38	14.87 ± 5.85	14.31 ± 4.27
ICP pulse morphology	10.95 ± 3.30	14.48 ± 5.04	14.86 ± 5.96	14.54 ± 4.95
History of target	$\mathbf{10.09 \pm 2.99}$	$\mathbf{11.98 \pm 3.85}$	$\mathbf{12.96 \pm 3.68}$	17.13 ± 5.21
SAX BoW	10.97 ± 3.30	13.75 ± 5.14	15.47 ± 6.29	14.79 ± 4.42
ICP-ABP trace	11.07 ± 3.24	13.41 ± 4.98	15.47 ± 6.29	14.79 ± 4.42

Table A.5: Average root mean squared error (RMSE) and standard deviation of all models and the baseline model when predicting Arterial Blood Pressure (ABP).

A.2 Cambridge

Feature Set	5 min	30 min	60 min	120 min
Baseline	5.76 ± 1.93	9.04 ± 2.59	10.60 ± 2.50	9.95 ± 1.91
All features	$\underline{5.62 \pm 1.88}$	8.75 ± 2.31	10.56 ± 2.48	9.80 ± 1.89
Statistical summaries	5.84 ± 1.78	8.88 ± 2.60	10.16 ± 2.46	9.81 ± 2.06
Ext. stat. summaries	5.62 ± 1.86	$\mathbf{8.27 \pm 1.99}$	$\underline{8.97 \pm 1.77}$	9.78 ± 1.99
Frequency features	5.80 ± 1.79	10.94 ± 7.58	9.70 ± 2.17	9.80 ± 1.99
CA indices	5.78 ± 1.86	8.86 ± 2.59	10.04 ± 2.42	9.81 ± 2.02
Ext. CA indices	$\mathbf{5.55 \pm 1.78}$	$\underline{8.30 \pm 2.18}$	$\mathbf{8.94 \pm 1.76}$	9.86 ± 1.83
Huser et al.	5.72 ± 1.96	8.85 ± 2.45	10.50 ± 2.14	$\mathbf{9.70 \pm 1.76}$
ICP pulse morphology	5.73 ± 1.87	8.87 ± 2.56	10.01 ± 2.41	$\underline{9.75 \pm 2.03}$
History of target	5.75 ± 1.49	8.71 ± 1.38	11.62 ± 3.72	16.32 ± 12.02
SAX BoW	5.73 ± 1.88	8.87 ± 2.55	10.43 ± 2.47	9.95 ± 1.91
ICP-ABP trace	5.88 ± 1.94	8.87 ± 2.72	10.69 ± 3.67	9.95 ± 1.91

Table A.6: Average mean absolute error (MAE) and standard deviation of all models and the baseline model when predicting Cerebral Perfusion Pressure (CPP).

Feature Set	5 min	30 min	60 min	120 min
Baseline	11.64 ± 3.37	15.11 ± 4.93	16.75 ± 5.23	15.00 ± 3.61
All features	10.88 ± 3.03	14.23 ± 4.23	16.53 ± 5.35	14.81 ± 3.48
Statistical summaries	11.33 ± 3.41	14.17 ± 4.64	15.64 ± 4.68	$\underline{14.53 \pm 3.64}$
Ext. stat. summaries	10.99 ± 3.43	$\mathbf{13.34 \pm 3.87}$	$\underline{14.00 \pm 3.53}$	14.56 ± 3.63
Frequency features	11.24 ± 3.33	17.67 ± 12.74	14.67 ± 3.85	14.54 ± 3.64
CA indices	11.40 ± 3.39	14.24 ± 4.65	15.56 ± 4.76	14.54 ± 3.70
Ext. CA indices	10.80 ± 3.41	13.48 ± 4.05	$\mathbf{13.94 \pm 3.50}$	14.64 ± 3.30
Huser et al.	$\underline{10.73 \pm 3.32}$	14.21 ± 4.91	16.35 ± 4.77	14.63 ± 2.98
ICP pulse morphology	11.35 ± 3.38	14.17 ± 4.61	15.57 ± 4.71	$\mathbf{14.50 \pm 3.56}$
History of target	$\mathbf{10.47 \pm 2.94}$	$\underline{13.43 \pm 2.84}$	15.95 ± 3.27	20.34 ± 13.08
SAX BoW	11.39 ± 3.31	14.35 ± 4.69	16.24 ± 4.81	15.00 ± 3.61
ICP-ABP trace	11.38 ± 3.35	14.16 ± 4.48	15.67 ± 5.20	15.00 ± 3.61

Table A.7: Average root mean squared error (RMSE) and standard deviation of all models and the baseline model when predicting Cerebral Perfusion Pressure (CPP).

Feature Set	5 min	30 min	60 min	120 min
Baseline	0.02 ± 0.01	0.04 ± 0.02	0.05 ± 0.02	0.05 ± 0.02
All features	**0.02 ± 0.01**	0.04 ± 0.01	0.05 ± 0.02	0.07 ± 0.03
Statistical summaries	0.02 ± 0.01	0.05 ± 0.02	0.05 ± 0.02	0.06 ± 0.04
Ext. stat. summaries	0.02 ± 0.01	0.04 ± 0.01	0.05 ± 0.02	0.05 ± 0.01
Frequency features	0.03 ± 0.01	0.05 ± 0.02	0.05 ± 0.02	0.05 ± 0.02
CA indices	0.02 ± 0.01	0.04 ± 0.02	0.05 ± 0.02	0.05 ± 0.01
Ext. CA indices	0.02 ± 0.01	0.05 ± 0.02	0.05 ± 0.02	0.06 ± 0.01
Huser et al.	0.02 ± 0.01	0.04 ± 0.01	0.06 ± 0.03	<u>0.05 ± 0.02</u>
ICP pulse morphology	0.02 ± 0.01	0.04 ± 0.02	**0.05 ± 0.02**	1304.71 ± 2917.31
History of target	0.02 ± 0.01	0.05 ± 0.03	0.05 ± 0.02	**0.05 ± 0.02**
SAX BoW	<u>0.02 ± 0.01</u>	**0.04 ± 0.02**	<u>0.05 ± 0.02</u>	0.05 ± 0.02
ICP-ABP trace	0.12 ± 0.22	<u>0.04 ± 0.02</u>	0.05 ± 0.01	0.05 ± 0.02

Table A.8: Average mean absolute error (MAE) and standard deviation of all models and the baseline model when predicting Transfer Function Index (TF).

Feature Set	5 min	30 min	60 min	120 min
Baseline	0.04 ± 0.01	0.06 ± 0.02	0.06 ± 0.02	0.07 ± 0.02
All features	0.04 ± 0.01	<u>0.06 ± 0.01</u>	**0.06 ± 0.02**	0.08 ± 0.03
Statistical summaries	0.04 ± 0.01	0.07 ± 0.02	0.07 ± 0.02	0.08 ± 0.04
Ext. stat. summaries	0.04 ± 0.01	**0.06 ± 0.01**	0.07 ± 0.02	0.07 ± 0.02
Frequency features	0.04 ± 0.01	0.07 ± 0.02	0.08 ± 0.03	0.07 ± 0.02
CA indices	0.04 ± 0.01	0.06 ± 0.02	0.07 ± 0.02	**0.07 ± 0.02**
Ext. CA indices	0.04 ± 0.01	0.07 ± 0.02	0.08 ± 0.03	0.08 ± 0.02
Huser et al.	0.04 ± 0.01	0.06 ± 0.01	0.08 ± 0.03	0.07 ± 0.02
ICP pulse morphology	**0.04 ± 0.01**	0.07 ± 0.02	<u>0.06 ± 0.02</u>	2536.68 ± 5672.04
History of target	0.04 ± 0.01	0.07 ± 0.02	0.07 ± 0.02	<u>0.07 ± 0.02</u>
SAX BoW	<u>0.04 ± 0.01</u>	0.06 ± 0.02	0.06 ± 0.02	0.07 ± 0.02
ICP-ABP trace	0.20 ± 0.36	0.06 ± 0.02	0.07 ± 0.02	0.07 ± 0.02

Table A.9: Average root mean squared error (RMSE) and standard deviation of all models and the baseline model when predicting Transfer Function Index (TF).

Feature Set	5 min	30 min	60 min	120 min
Baseline	0.17 ± 0.04	0.28 ± 0.07	0.25 ± 0.06	0.30 ± 0.11
All features	$\underline{0.17 \pm 0.04}$	0.28 ± 0.06	0.25 ± 0.06	0.30 ± 0.11
Statistical summaries	0.17 ± 0.04	0.51 ± 0.25	0.48 ± 0.37	0.31 ± 0.10
Ext. stat. summaries	0.20 ± 0.11	0.29 ± 0.06	0.25 ± 0.06	$\underline{0.29 \pm 0.07}$
Frequency features	0.17 ± 0.04	0.28 ± 0.07	0.25 ± 0.06	0.41 ± 0.32
CA indices	0.17 ± 0.04	2.98 ± 2.52	0.31 ± 0.06	0.54 ± 0.34
Ext. CA indices	0.17 ± 0.04	0.31 ± 0.06	0.25 ± 0.06	0.30 ± 0.10
Huser et al.	0.28 ± 0.27	0.28 ± 0.07	0.25 ± 0.06	0.30 ± 0.11
ICP pulse morphology	0.19 ± 0.04	0.65 ± 0.58	0.38 ± 0.22	0.36 ± 0.09
History of target	$\mathbf{0.16 \pm 0.03}$	$\mathbf{0.26 \pm 0.04}$	$\mathbf{0.25 \pm 0.05}$	$\mathbf{0.26 \pm 0.07}$
SAX BoW	0.17 ± 0.04	$\underline{0.28 \pm 0.07}$	0.25 ± 0.06	0.29 ± 0.12
ICP-ABP trace	0.17 ± 0.04	0.28 ± 0.07	$\underline{0.25 \pm 0.06}$	0.30 ± 0.11

Table A.10: Average mean absolute error (MAE) and standard deviation of all models and the baseline model when predicting Single Wave ICP-ABP Amplitude Correlation (IAAC).

Feature Set	5 min	30 min	60 min	120 min
Baseline	0.25 ± 0.07	0.38 ± 0.12	0.34 ± 0.10	0.39 ± 0.15
All features	$\underline{0.25 \pm 0.07}$	0.38 ± 0.11	0.34 ± 0.10	0.39 ± 0.15
Statistical summaries	0.25 ± 0.07	0.84 ± 0.51	0.72 ± 0.71	0.41 ± 0.16
Ext. stat. summaries	0.29 ± 0.12	0.40 ± 0.11	0.34 ± 0.11	$\underline{0.37 \pm 0.09}$
Frequency features	0.25 ± 0.07	0.38 ± 0.11	0.34 ± 0.11	0.49 ± 0.33
CA indices	0.25 ± 0.07	6.64 ± 7.34	0.40 ± 0.09	0.78 ± 0.58
Ext. CA indices	0.25 ± 0.07	0.43 ± 0.10	0.34 ± 0.11	0.39 ± 0.15
Huser et al.	0.36 ± 0.27	0.38 ± 0.12	0.34 ± 0.11	0.39 ± 0.15
ICP pulse morphology	0.28 ± 0.07	1.19 ± 1.11	0.54 ± 0.38	0.50 ± 0.22
History of target	$\mathbf{0.24 \pm 0.07}$	$\mathbf{0.33 \pm 0.06}$	$\mathbf{0.31 \pm 0.06}$	$\mathbf{0.33 \pm 0.09}$
SAX BoW	0.25 ± 0.07	$\underline{0.38 \pm 0.12}$	0.34 ± 0.10	0.39 ± 0.16
ICP-ABP trace	0.25 ± 0.07	0.38 ± 0.12	$\underline{0.34 \pm 0.11}$	0.39 ± 0.15

Table A.11: Average root mean squared error (RMSE) and standard deviation of all models and the baseline model when predicting Single Wave ICP-ABP Amplitude Correlation (IAAC).

Record	Samples	ABP	CPP	IAAC	ICP	PRx	TF	Length
0	2436	96%	96%	81%	96%	93%	93%	0 days 06:46:00
1	4965	98%	98%	98%	98%	98%	98%	0 days 13:47:30
2	2592	95%	95%	96%	99%	97%	97%	0 days 07:12:00
3	3355	50%	50%	42%	50%	49%	49%	0 days 09:19:10
8	4193	77%	77%	76%	78%	76%	76%	0 days 11:38:50

Table A.12: Availability of the individual targets at the *5 minutes* prediction horizon for each record from the Cambridge data set.

Record	Samples	ABP	CPP	IAAC	ICP	PRx	TF	Length
0	2436	89%	89%	70%	89%	86%	86%	0 days 06:46:00
1	4965	95%	95%	95%	95%	95%	95%	0 days 13:47:30
2	2592	89%	89%	91%	93%	91%	91%	0 days 07:12:00
3	3355	46%	46%	33%	46%	45%	45%	0 days 09:19:10
8	4193	73%	73%	72%	75%	72%	72%	0 days 11:38:50

Table A.13: Availability of the individual targets at the *30 minutes* prediction horizon for each record from the Cambridge data set.

Record	Samples	ABP	CPP	IAAC	ICP	PRx	TF	Length
0	2436	67%	67%	51%	67%	64%	64%	0 days 06:46:00
1	4965	84%	84%	84%	84%	84%	84%	0 days 13:47:30
2	2592	70%	70%	70%	72%	70%	70%	0 days 07:12:00
3	3355	30%	30%	18%	30%	29%	29%	0 days 09:19:10
8	4193	60%	60%	59%	62%	59%	59%	0 days 11:38:50

Table A.14: Availability of the individual targets at the *120 minutes* prediction horizon for each record from the Cambridge data set.

A.3 Performance Evaluation

We did a small performance evaluation to measure the achievable speed up when using compiled code for feature construction. For this we set up a micro benchmark evaluating the runtime of the original function and the runtime of the compiled function on input vectors of approximately the length used in the framework. We did three warmup rounds to fill the cache and JIT compile the functions. Then we evaluated each function at least 5 times and at least for 1 second, whatever took longer. The results can be seen in Figure A.9. We only list the functions were we actually achieved a performance improvement. We show the slow down factor of the interpreted code compared to the compiled code. Each subfigure shows the result for one function. One the left is the original Python implementation, on the right is the function either implemented in Python and JIT compiled or im-

plemented in Cython. We also annotated each title with the mean runtime of the compiled function to show each functions absolute runtime. These can then be put in context with the actual feature construction time for one window, which is approximately 9 ms. Thus, not all of the benchmarked functions were actually used to construct the feature set of this work.

In many functions we could achieve a speed up of at least 25, in some even a factor of 100. The one exception is the QRS location algorithm where we only achieved a speed up of two. We assume that this is because we still need to call back into the python runtime multiple times inside the function and the code cannot be maximally optimized.

We choosing the subsegment for benchmarking we took care that the functions actually need to do work. We selected a segment, where QRS detection algorithms would find pulses and where there was no missing value.

We also compared the runtime of a FIR filter (Kaiser) with an IIR filter (Butterworth) to show the achievable speed up when switching the type of the filter. Both were low-pass filters with a cutoff frequency of 15 Hz for ICP and 20 Hz for ABP. To construct the Kaiser filter we specified a width of 0.5 Hz and a max ripple of 60 db which were both the parameters used in the framework. The main issue with the FIR filter was, that it had over a thousand coefficients compared to just 6 coefficients of the IIR filter. Figure A.10 show the summary of the two filter types. The IIR filter is clearly at least a factor of 20 faster than the FIR filter. To put the absolute numbers into context we also annotated the title with the absolute run time of the IIR filter.
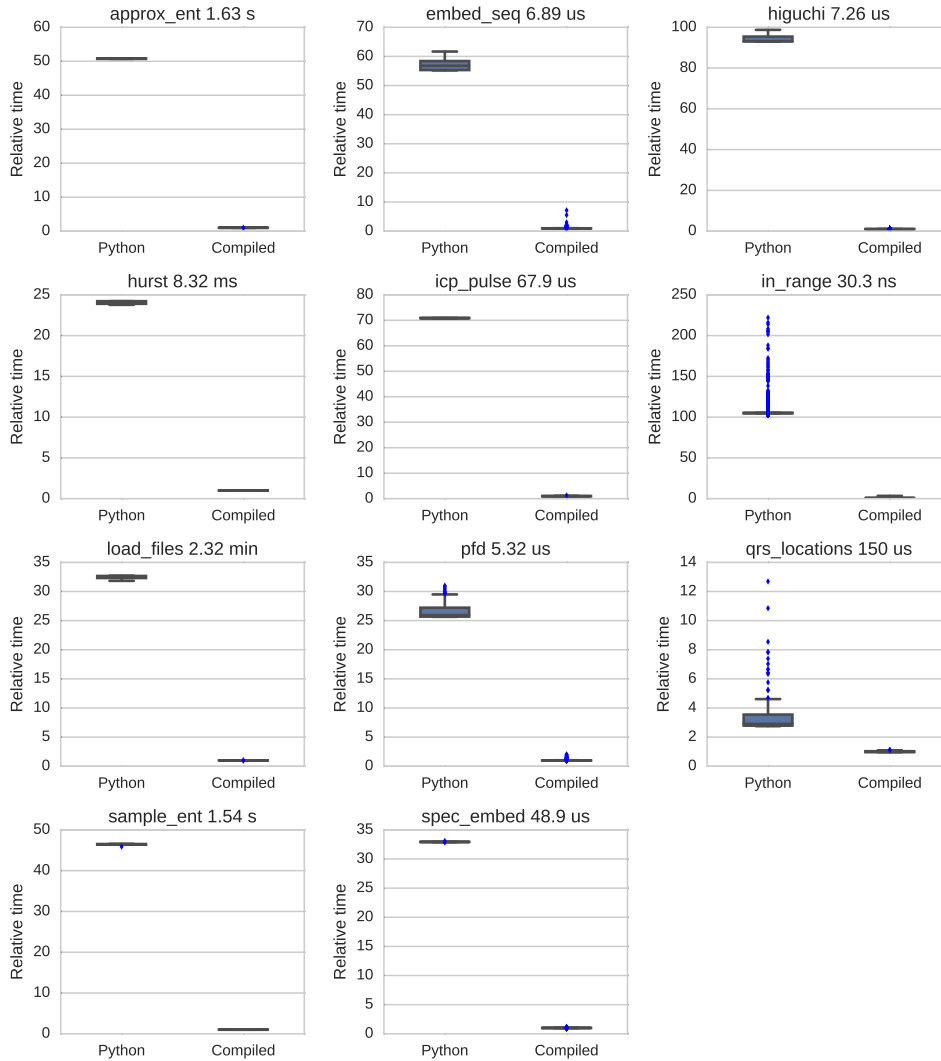
Figure A.9: Achieved performance improvement when using (JIT) compiled feature construction functions. Here we list the 11 functions were we were able to achieve a speed up. We normalized the time values to the mean runtime of the compiled function to see the relative speed up.
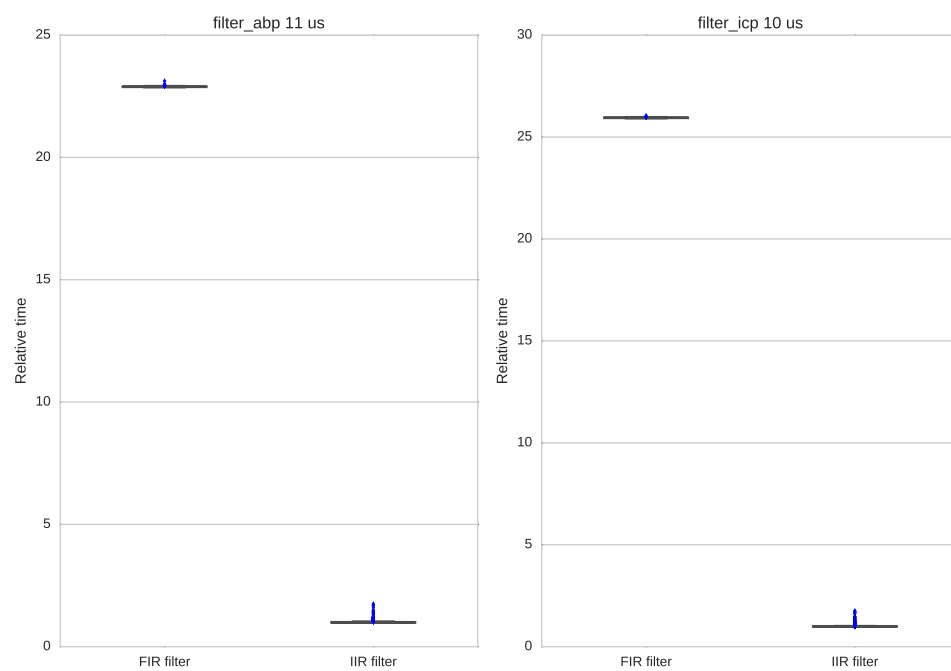
Figure A.10: Achieved performance improvement when using an IIR filter instead of an FIR filter at a cutoff of 15 Hz.
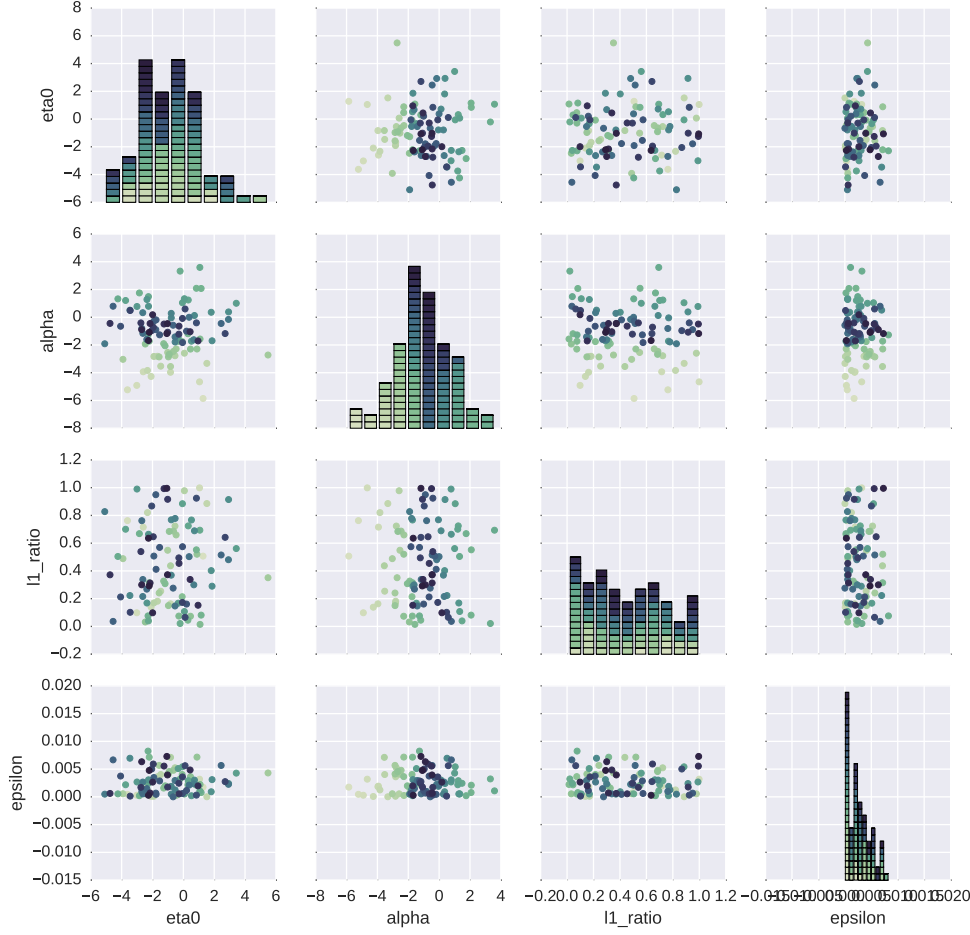
Figure A.11: The results of a hyperparameter search for the prediction target PRx at prediction horizon 30 minutes for the feature set of CA indices. A lower cross-validation error is denoted by a darker color of the point. The pairwise scatter plots show the interactions between the four hyperparameters. The parameter eta0 denotes the initial step size and is log-normal distributed. Thus we report the logarithmic scale. The parameter alpha denotes the regularization coefficient. Since it is also log-normal distributed we also report the logarithmic scale. The parameter l1_ratio denotes the ratio in [0..1] of the l1 loss in the total regularization loss of l1 + l2. Since it is uniform distributed we report the linear scale. The parameter epsilon denotes the width of the epsilon-insensitive region of the absolute loss wherein the loss gets truncated to zero. Since it is exponentially distributed we also report the log-scale. For all parameters we see that the search algorithm covered the parameter space well such that there are no one-sided darker distributions.

Figure A.12: A representative segment of ICP comparing the actual prediction target with the baseline prediction and the prediction of the best model. The baseline prediction also represents the current value of the prediction target at that point in time.
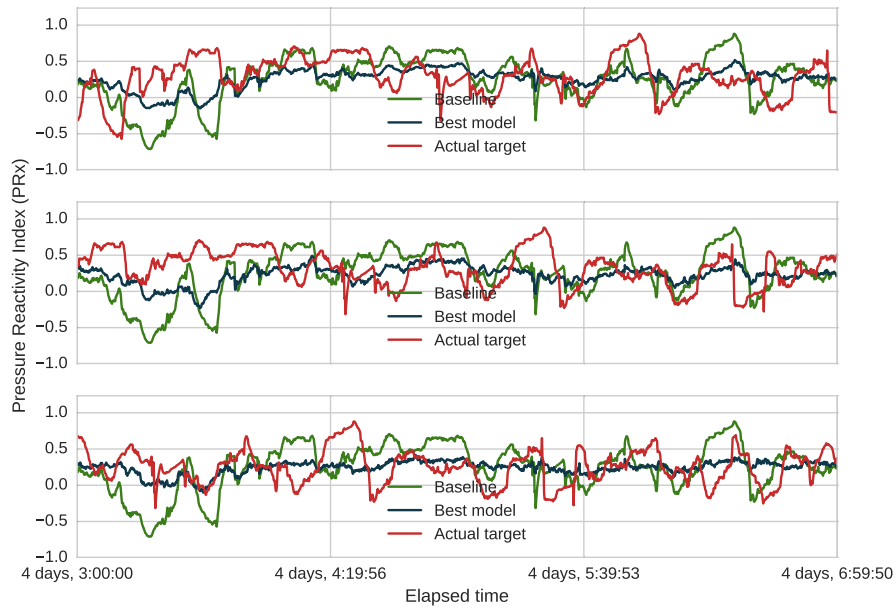
Figure A.13: A representative segment of PRx comparing the actual prediction target with the baseline prediction and the prediction of the best model. The baseline prediction also represents the current value of the prediction target at that point in time.

Figure A.14: A representative segment of IAAC comparing the actual prediction target with the baseline prediction and the prediction of the best model. The baseline prediction also represents the current value of the prediction target at that point in time.

Bibliography

[1] Marcel JH Aries, Marek Czosnyka, Karol P Budohoski, Luzius A Steiner, Andrea Lavinio, Angelos G Kolias, Peter J Hutchinson, Ken M Brady, David K Menon, John D Pickard, et al. Continuous determination of optimal cerebral perfusion pressure in traumatic brain injury*. *Critical care medicine*, 40(8):2456–2463, 2012.

[2] Marcel JH Aries, Jan W Elting, Jacques De Keyser, Berry PH Kremer, and Patrick CAJ Vroomen. Cerebral autoregulation in stroke a review of transcranial doppler studies. *Stroke*, 41(11):2697–2704, 2010.

[3] CJ Avezaat, JH Van Eijndhoven, and DJ Wyper. Cerebrospinal fluid pulse pressure and intracranial volume-pressure relationships. *Journal of Neurology, Neurosurgery & Psychiatry*, 42(8):687–700, 1979.

[4] Murad Banaji, Ilias Tachtsidis, David Delpy, and Stephen Baigent. A physiological model of cerebral blood flow control. *Mathematical biosciences*, 194(2):125–173, 2005.

[5] Yoshua Bengio. Practical recommendations for gradient-based training of deep architectures. In *Neural Networks: Tricks of the Trade*, pages 437–478. Springer, 2012.

[6] Léon Bottou. Stochastic gradient descent tricks. In *Neural Networks: Tricks of the Trade*, pages 421–436. Springer, 2012.

[7] SL Bratton, RM Chestnut, J Ghajar, Hammond FF McConnell, OA Harris, R Hartl, GT Manley, A Nemecek, DW Newell, G Rosenthal, et al. Guidelines for the management of severe traumatic brain injury. viii. intracranial pressure thresholds. *Journal of neurotrauma*, 24:S55–8, 2006.

[8] Karol P Budohoski, Marek Czosnyka, Peter J Kirkpatrick, Peter Smielewski, Luzius A Steiner, and John D Pickard. Clinical relevance

of cerebral autoregulation following subarachnoid haemorrhage. *Nature Reviews Neurology*, 9(3):152–163, 2013.

[9] Oliver G Cameron, Jack G Modell, and M Hariharan. Caffeine and human cerebral blood flow: a positron emission tomography study. *Life sciences*, 47(13):1141–1146, 1990.

[10] JM Chillon and GL Baumbach. Autoregulation: arterial and intracranial pressure. *Cerebral blood flow and metabolism*, 2:395–412, 2002.

[11] M Czosnyka, P Smielewski, P Kirk-patrick, DK Menon, and J D Pickard. Monitoring of cerebral autoregulation in head-injured patients. *Journal of Neurosurgical Anesthesiology*, 9(2):200, 1997.

[12] Marek Czosnyka and John D Pickard. Monitoring and interpretation of intracranial pressure. *Journal of Neurology, Neurosurgery & Psychiatry*, 75(6):813–821, 2004.

[13] Marek Czosnyka, Piotr Smielewski, Peter Kirkpatrick, Rodney J Laing, David Menon, and John D Pickard. Continuous assessment of the cerebral vasomotor reactivity in head injury. *Neurosurgery*, 41(1):11–19, 1997.

[14] Ingrid Daubechies. The wavelet transform, time-frequency localization and signal analysis. *Information Theory, IEEE Transactions on*, 36(5):961–1005, 1990.

[15] Rolf R Diehl, Dieter Linden, Dorothee Lücke, and Peter Berlit. Phase relationship between cerebral blood flow velocity and blood pressure a clinical test of autoregulation. *Stroke*, 26(10):1801–1804, 1995.

[16] Rolf R Diehl, Dieter Linden, Dorothee Lücke, and Peter Berlit. Spontaneous blood pressure oscillations and cerebral autoregulation. *Clinical Autonomic Research*, 8(1):7–12, 1998.

[17] Per Kristian Eide. A new method for processing of continuous intracranial pressure signals. *Medical engineering & physics*, 28(6):579–587, 2006.

[18] Per Kristian Eide, Gunnar Bentsen, Angelika G Sorteberg, Pål Bache Marthinsen, Audun Stubhaug, and Wilhelm Sorteberg. A randomized and blinded single-center trial comparing the effect of intracranial pressure and intracranial pressure wave amplitude-guided intensive care management on early clinical state and 12-month outcome in patients with aneurysmal subarachnoid hemorrhage. *Neurosurgery*, 69(5):1105–1115, 2011.

[19] Per Kristian Eide, Angelika Sorteberg, Gunnar Bentsen, Pål Bache Marthinsen, Audun Stubhaug, and Wilhelm Sorteberg. Pressure-derived versus pressure wave amplitude–derived indices of cerebrovascular pressure reactivity in relation to early clinical state and 12-month outcome following aneurysmal subarachnoid hemorrhage: Clinical article. *Journal of neurosurgery*, 116(5):961–971, 2012.

[20] PK Eide, E-H Park, and JR Madsen. Arterial blood pressure vs intracranial pressure in normal pressure hydrocephalus. *Acta neurologica Scandinavica*, 122(4):262–269, 2010.

[21] Charles D Fraser, Ken M Brady, Christopher J Rhee, R Blaine Easley, Kathleen Kibler, Peter Smielewski, Marek Czosnyka, David W Kaczka, Dean B Andropoulos, and Craig Rusin. The frequency response of cerebral autoregulation. *Journal of Applied Physiology*, 115(1):52–56, 2013.

[22] Erzhen Gao, William L Young, John Pile-Spellman, Eugene Ornstein, and Qiyuan Ma. Mathematical considerations for modeling cerebral blood flow autoregulation to systemic arterial pressure. *American Journal of Physiology-Heart and Circulatory Physiology*, 274(3):H1023–H1031, 1998.

[23] Cole A Giller, Gary Bowman, Hunter Dyer, Lee Mootz, and William Krippner. Cerebral arterial diameters during changes in blood pressure and carbon dioxide during craniotomy. *Neurosurgery*, 32(5):737–742, 1993.

[24] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.

[25] Fabian Güiza, Bart Depreitere, Ian Piper, Greet Van den Berghe, and Geert Meyfroidt. Novel methods to predict increased intracranial pressure during intensive care and long-term neurologic outcome after traumatic brain injury: Development and validation in a multicenter dataset*. *Critical care medicine*, 41(2):554–564, 2013.

[26] Robert Hamilton, Peng Xu, Shadnaz Asgari, Magdalena Kasprowicz, Paul Vespa, Marvin Bergsneider, and Xaio Hu. Forecasting intracranial pressure elevation using pulse waveform morphology. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 4331–4334. IEEE, 2009.

[27] Xiao Hu, Peng Xu, Shadnaz Asgari, Paul Vespa, and Marvin Bergsneider. Forecasting icp elevation based on prescient changes of intracranial pressure waveform morphology. *Biomedical Engineering, IEEE Transactions on*, 57(5):1070–1078, 2010.

[28] Matthias Hüser. Forecasting intracranial hypertension using time series and waveform features. Master's thesis, ETH Zürich, Switzerland, 2015.

[29] Matthias Hüser, Valeria De Luca, Martin Jaggi, Walter Karlen, and Emanuela Keller. Forecasting intracranial hypertension using waveform and time series features. Vasospasm 2015 - 13th International Conference on Neurovascular Events after Subarachnoid Hemorrhage Program, 2015.

[30] Faisal M Kashif, George C Verghese, Vera Novak, Marek Czosnyka, and Thomas Heldt. Model-based noninvasive estimation of intracranial pressure from cerebral blood flow velocity and arterial pressure. *Science translational medicine*, 4(129):129ra44–129ra44, 2012.

[31] Dong-Joo Kim, Zofia Czosnyka, Nicole Keong, Danila K Radolovich, Peter Smielewski, Michael PF Sutcliffe, John D Pickard, and Marek Czosnyka. Index of cerebrospinal compensatory reserve in hydrocephalus. *Neurosurgery*, 64(3):494–502, 2009.

[32] Niels A Lassen. *Cerebral blood flow and oxygen consumption in man*. Am Physiological Soc, 1959.

[33] Miroslaw Latka, Malgorzata Turalska, Marta Glaubic-Latka, Waldemar Kolodziej, Dariusz Latka, and Bruce J West. Phase dynamics in cerebral autoregulation. *American Journal of Physiology-Heart and Circulatory Physiology*, 289(5):H2272–H2279, 2005.

[34] Christos Lazaridis, Stacia M DeSantis, Peter Smielewski, David K Menon, Peter Hutchinson, John D Pickard, and Marek Czosnyka. Patient-specific thresholds of intracranial pressure in severe traumatic brain injury: Clinical article. *Journal of neurosurgery*, 120(4):893–900, 2014.

[35] Christos Lazaridis, Piotr Smielewski, Luzius A Steiner, Ken M Brady, Peter Hutchinson, John D Pickard, and Marek Czosnyka. Optimal cerebral perfusion pressure: are we ready for it? *Neurological research*, 35(2):138–148, 2013.

[36] JJ Lemaire, T Khalil, F Cervenansky, G Gindre, JY Boire, JE Bazin, B Irthum, and J Chazal. Slow pressure waves in the cranial enclosure. *Acta neurochirurgica*, 144(3):243–254, 2002.

[37] Jessica Lin, Eamonn Keogh, Li Wei, and Stefano Lonardi. Experiencing sax: a novel symbolic representation of time series. *Data Mining and knowledge discovery*, 15(2):107–144, 2007.

[38] Jessica Lin and Yuan Li. Finding structural similarity in time series data using bag-of-patterns representation. In *Scientific and Statistical Database Management*, pages 461–477. Springer, 2009.

[39] Georgios D Mitsis, Marc J Poulin, Peter A Robbins, and Vasilis Z Marmarelis. Nonlinear modeling of the dynamic effects of arterial pressure and co 2 variations on cerebral blood flow in healthy humans. *Biomedical Engineering, IEEE Transactions on*, 51(11):1932–1943, 2004.

[40] David W Newell, Rune Aaslid, Arthur Lam, Teresa S Mayberg, and H Richard Winn. Comparison of flow and velocity during dynamic autoregulation testing in humans. *Stroke*, 25(4):793–797, 1994.

[41] Ronney B Panerai. Assessment of cerebral pressure autoregulation in humans-a review of measurement methods. *Physiological measurement*, 19(3):305, 1998.

[42] Gianfranco Parati, Paolo Castiglioni, Marco Di Rienzo, Stefano Omboni, Antonio Pedotti, and Giuseppe Mancia. Sequential spectral analysis of 24-hour blood pressure and pulse interval in humans. *Hypertension*, 16(4):414–421, 1990.

[43] Gianfranco Parati, J Philip Saul, Marco Di Rienzo, and Giuseppe Mancia. Spectral analysis of blood pressure and heart rate variability in evaluating cardiovascular regulation a critical appraisal. *Hypertension*, 25(6):1276–1286, 1995.

[44] Joseph E Parrillo and R Phillip Dellinger. *Critical Care Medicine: Principles of Diagnosis and Management in the Adult (Expert Consult-Online and Print)*. Elsevier Health Sciences, 2013.

[45] OB Paulson, S Strandgaard, and L Edvinsson. Cerebral autoregulation. *Cerebrovascular and brain metabolism reviews*, 2(2):161–192, 1989.

[46] Marc J Poulin and Peter A Robbins. Indexes of flow and cross-sectional area of the middle cerebral artery using doppler ultrasound during hypoxia and hypercapnia in humans. *Stroke*, 27(12):2244–2250, 1996.

[47] DK Radolovich, MJH Aries, G Castellani, A Corona, A Lavinio, P Smielewski, JD Pickard, and M Czosnyka. Pulsatile intracranial pressure and cerebral autoregulation after traumatic brain injury. *Neurocritical care*, 15(3):379–386, 2011.

[48] Matthias Reinhard, Andreas Hetzel, Michael Lauk, and Carl H Lücking. Dynamic cerebral autoregulation testing as a diagnostic tool in patients with carotid artery stenosis. *Neurological research*, 23(1):55–63, 2001.

[49] Mohammed Saeed, Mauricio Villarroel, Andrew T Reisner, Gari Clifford, Li-Wei Lehman, George Moody, Thomas Heldt, Tin H Kyaw, Benjamin Moody, and Roger G Mark. Multiparameter intelligent monitoring in intensive care ii (mimic-ii): a public-access intensive care unit database. *Critical care medicine*, 39(5):952, 2011.

[50] Werner Schregel, Heinrich Schaefermeyer, Marian Sihle-Wissel, and Rebecca Klein. Transcranial doppler sonography during isoflurane/n2o anaesthesia and surgery: flow velocity,"vessel area" and "volume flow". *Canadian journal of anaesthesia*, 41(7):607–612, 1994.

[51] Shahid Shafi, Ramon Diaz-Arrastia, Christopher Madden, and Larry Gentilello. Intracranial pressure monitoring in brain-injured patients is associated with worsening of survival. *Journal of Trauma and Acute Care Surgery*, 64(2):335–340, 2008.

[52] Shai Shalev-Shwartz, Yoram Singer, Nathan Srebro, and Andrew Cotter. Pegasos: Primal estimated sub-gradient solver for svm. *Mathematical programming*, 127(1):3–30, 2011.

[53] Luzius A Steiner, Marek Czosnyka, Stefan K Piechnik, Piotr Smielewski, Doris Chatfield, David K Menon, and John D Pickard. Continuous monitoring of cerebrovascular pressure reactivity allows determination of optimal cerebral perfusion pressure in patients with traumatic brain injury. *Critical care medicine*, 30(4):733–738, 2002.

[54] K Tsutsumi, K Ueki, M Usui, S Kwak, and T Kirino. Risk of subarachnoid hemorrhage after surgical treatment of unruptured cerebral aneurysms. *Stroke*, 30(6):1181–1184, 1999.

[55] Arenda HEA Van Beek, Jurgen AHR Claassen, Marcel GM Olde Rikkert, and René WMM Jansen. Cerebral autoregulation: an overview of current concepts and methodology with special focus on the elderly. *Journal of Cerebral Blood Flow & Metabolism*, 28(6):1071–1085, 2008.

[56] Feng Zhang, Mengling Feng, Sinno Jialin Pan, Liang Yu Loy, Wenyuan Guo, Zhuo Zhang, Pei Loon Chin, Cuntai Guan, Nicolas Kon Kam King, and Beng Ti Ang. Artificial neural network based intracranial pressure mean forecast algorithm for medical decision support. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 7111–7114. IEEE, 2011.

[57] Rong Zhang, Julie H Zuckerman, Cole A Giller, and Benjamin D Levine. Transfer function analysis of dynamic cerebral autoregulation in humans. *American Journal of Physiology-Heart and Circulatory Physiology*, 274(1):H233–H241, 1998.

[58] Rong Zhang, Julie H Zuckerman, Kenichi Iwasaki, Thad E Wilson, Craig G Crandall, and Benjamin D Levine. Autonomic neural control of dynamic cerebral autoregulation in humans. *Circulation*, 106(14):1814–1820, 2002.

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

Prediction of Cerebral Autoregulation in Intensive Care Patients

Authored by (in block letters):
For papers written by groups the names of all authors are required.

Name(s):	First name(s):
Kündig	Adrian

With my signature I confirm that
- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date	Signature(s)
Zürich, 26.01.2016	*[signature]*

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.