

Predicting intracranial hypertension in acute traumatic brain injury patients

Master thesis, Medical Sensing and Stimulation

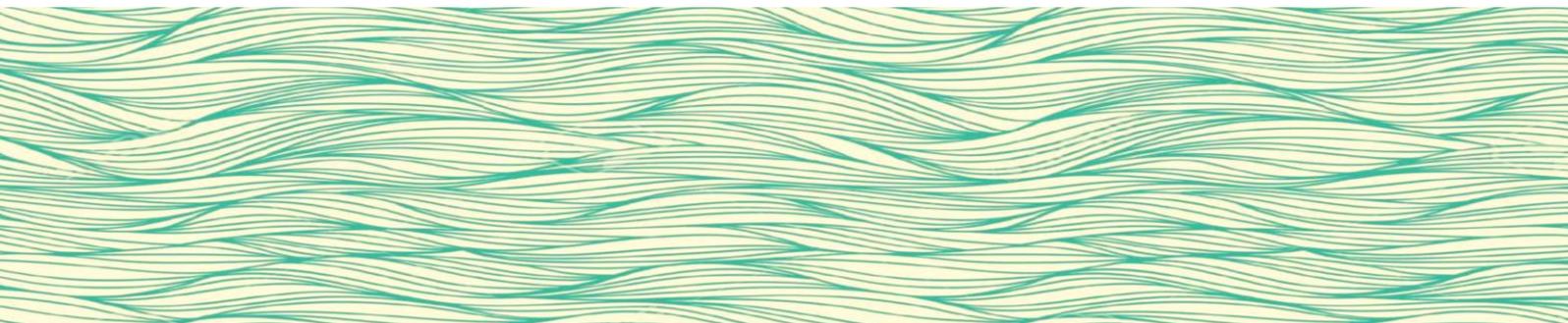
Technical Medicine

Jantine Smit

Supervisors:

Chairman and technical supervisor:	Prof. dr. ir. M.J.A.M. van Putten
Medical supervisor:	dr. C.W.E. Hoedemaekers
Technical supervisor institution:	C.R. van Kaam, MSc
Technical supervisor:	dr. M.C. Tjepkema-Cloostermans
External supervisor:	dr. ir. B.E. Westerhof
Process supervisor:	drs. E.M. Walter

UNIVERSITY OF TWENTE.
Radboudumc



Acknowledgements

This thesis represents the finish line of my master in Medical Sensing and Stimulation. Over the last six years I have had the great opportunity of studying Technical Medicine at the University of Twente, it represents a period I will never forget.

I would like to thank my master thesis supervisors Astrid Hoedemaekers, Ruud van Kaam, Michel van Putten, Marleen Tjepkema – Cloostermans and Elyse Walter. Thank you for the feedback, support and discussions throughout my internship. I am grateful to the Radboud UMC, Intensive Care for hosting me and for all the clinical opportunities.

Finally, I would like to thank my fellow students, friends, family and Tim for always being there for me!

Jantine Smit

Nijmegen, February 2020

Abstract

Objective: Prediction of intracranial hypertension in acute traumatic brain injury patients. We studied the potential to predict intracranial hypertension by morphological features of the ICP signal and by the use of a neural network.

Methods: Fifty-three traumatic brain injury patients were included in this retrospective study. These patients were admitted to the ICU at RadboudUMC, Nijmegen. The ICP was monitored using an intraparenchymal probe. Spectral regression analysis was performed to robustly detect the ICP-subpeaks and calculate 27 morphological metrics of the ICP signal (MOCAIP). The spectral regression method required a test and training set. 7 experts of the RadboudUMC labelled subpeaks in averaged ICP waves. The performance and consistency in peak labelling of the experts was evaluated by the percentage of inconsistency and intra-and-intraclass correlation coefficients. By lack of a gold standard the labelled waves were used as a gold standard to train and evaluate the performance of the spectral regression algorithm. A Kruskal Wallis test was performed to test for significance in the MOCAIP metrics between control waves, waves leading to intracranial hypertension (pre-IH) and intracranial hypertension (IH) waves. A classic machine learning approach was used to test the potential to predict intracranial hypertension based on the MOCAIP metrics by a classification tree. A CNN and LSTM neural network were used to classify IH and no intracranial hypertension (no-IH) based on the original ICP-waves.

Results: We found statistical significance in 11 metrics between the IH waves and pre-IH waves. 5 metrics were significant between ICP waves leading to hypertension and control waves. Waves leading to hypertension could be distinguished from control waves with a sensitivity of 89% and specificity of 71%. The predictive power in classification of pre-IH waves at specific timing prior to IH (5, 10, 15, 20 minutes) based on MOCAIP metrics is still limited. The obtained results from the neural network showed an accuracy of 70%, 97% sensitivity and 66 % specificity to classify between IH and no-IH waves. These results were very disparate for the test and training set.

Conclusion: Prediction of intracranial hypertension is promising for classification between ICP waves leading to hypertension and control waves based on the MOCAIP metrics.

List of abbreviations

TBI	Traumatic Brain Injury
ICU	Intensive care unit
IH	Intracranial hypertension
CPP	Cerebral perfusion pressure
ABP	Arterial blood pressure
ICP	Intracranial pressure
Pto2	Partial oxygen pressure
EEG	Electroencephalography
CSF	Cerebrospinal fluid
dV	Delta Volume
dP	Delta Pressure
CBF	Cerebral blood flow
MAP	Mean arterial pressure
MOCAIP	Morphological Analysis and Clustering of intracranial pressure
ICC	Intraclass correlation
KSR	Kernel Spectral Regression
LSTM	Long Short Term Memory
CNN	Convolutional Neural Network
SMI	Single Measure Intraclass Correlation

Table of contents

Acknowledgements	ii
Abstract	iv
List of abbreviations	vi
Table of contents	viii
Introduction	2
Method	6
Dataset.....	6
Preprocessing	6
Moving average filter to detect heart beat.....	6
Pulse preprocessing	7
Pulse averaging and normalization	7
Sub-peak detection as a regression problem.....	7
Expert labelling	8
Intraclass Correlation and inconsistent labelling	8
Spectral regression.....	9
Train algorithm	10
Detect candidate peaks	11
Sub-peaks.....	11
Position of sub-Peaks.....	11
Prediction accuracy	11
MOCAIP metrics.....	12
Segment selection	13
Classification tree - Morphological features	15
Statistical analysis of MOCAIP metrics	15
Classification tree	15
Neural Network – Averaged wave.....	15
Results	18
Dataset.....	18
Sub-peak detection as a regression problem.....	18
Expert Labelling - Inter-intraclass correlation and inconsistent labelling.....	18
Spectral regression.....	19
Peak prediction	19
MOCAIP metrics.....	22

Statistical analysis of MOCAIP metrics	22
Classification tree –Morphological features.....	24
Neural Network – Averaged wave.....	28
Discussion.....	30
Preprocessing	30
Expert labelling.....	30
Spectral Regression.....	31
MOCAIP.....	32
Classification by a classification tree - MOCAIP.....	33
Classification by a Neural Network.....	34
Clinical application	34
Conclusion	36
References	38
Appendices	42
Appendix 1 – Spectral Regression Algorithm.....	42
Appendix 2 – MOCAIP metrics.....	44
Appendix 3 – Interrater correlation coefficient and percentage of inconsistent labelling.....	46
Appendix 4 – Performance of spectral regression algorithm	48
Appendix 5 – MOCAIP statistics	50

Introduction

Traumatic brain injury (TBI), is a significant cause of mortality and morbidity in both children and adults [1,2]. Road traffic accidents and falls are the most common cause of TBI [3,4]. The focus in the critical care management of TBI patients is identification, prevention and treatment of secondary brain injury [5]. To supplement the clinical examination of TBI patients a combination of monitors, integrating physiologic and biological variables into assessment of the brain metabolism, oxygenation and perfusion is often used [6]. The combined use of multiple monitors is termed multimodal monitoring. Common parameters in multimodality monitoring are arterial blood pressure (ABP), intracranial pressure (ICP), partial oxygen pressure in the brain (Pto₂), electroencephalography (EEG) and transcranial doppler [1]. Bedside analysis and integration of these parameters has potential to provide more insight in the health-status of the patient and optimization of management of brain injured patients [6]. The role of multimodal monitoring in neurocritical care is still to be fully elucidated as it is rapidly evolving [5].

ICP monitoring is fundamental in the care of patients with TBI and is routinely used. The purpose of ICP monitoring is to control the intracranial pressure and maintain adequate cerebral perfusion pressure (CPP). Normal ICP in a supine healthy adult typically ranges from 7 to 15 mmHg [7]. Current guidelines suggest treating ICP greater than 20mmHg. Intracranial hypertension (IH) (ICP > 20 mmHg) poses a risk for secondary brain injury as it may impede the cerebral blood flow (CBF) and cause ischemia [8]. The causes of intracranial hypertension are diverse, examples are brain edema a growing intracranial mass or bleeding. ICP monitoring has been included in every guideline of the Brain Trauma Foundation for severe TBI management [9]. Relying solely on mean ICP is an accurate way to detect ICP elevation, but it carries a significant risk of treatment delay [10]. ICP treatment based on a numerical threshold may be oversimplification of the situation, as hypoxia or cellular dysfunction can occur in the 'normal' ICP range [11].

Many studies support the use of ICP monitoring. Most studies were observational and found that IH led to poorer outcomes, they concluded a significant lower mortality for patients treated using an ICP monitor compared with patients treated without an ICP monitor [12,13,14].

In the last decennia, numerous studies have tried to predict IH using time-series analysis and machine learning techniques [15]. They conducted research to automatically analyze ICP on the basis of mathematical models [16,17,18]. The collection and elaboration of large amounts of data on ICP waveform have therefore been possible, but the clinical value of these data is still a matter of debate [19]. Research showed that the onset of an intracranial pressure crisis can be predicted 30 minutes before IH with an area under the receiver operating characteristics curve of 0.86

[20,21]. This suggests the use of ICP data to predict long term neurological outcomes and acute secondary injury, further research is required to assess the impact of these predictions in clinical practice [21].

Changes in the ICP waveform might be informative for an incoming or established increase in the intracranial pressure[19]. The phenomenon of ICP pulse transiting from a normal three-peak configuration to a more rounded form is included in numerous machine learning methods. Some methods have been conducted to understand the characteristics in waveform morphology linked to elevated intracranial pressure [22,23,24,25]. Results obtained by Xiao Hu et al., using the Morphological Clustering and Analysis of Intracranial Pressure (MOCAIP) algorithm, indicated that an extensive set of ICP pulse morphological metrics may provide necessary information for forecasting ICP elevation. However, a heterogeneous group of patients was included in this study and the applicability of their algorithm on data of TBI patients is unknown [10]. Machine learning algorithms have demonstrated to be a viable approach for intracranial hypertension detection. Accessibility of these algorithms and MOCAIP methods remains an issue as these models are not easily understood by bedside clinicians, limiting their practical use. An algorithm and program which can be easily understood would be groundbreaking and enable patients to be treated in a proactive manner [25].

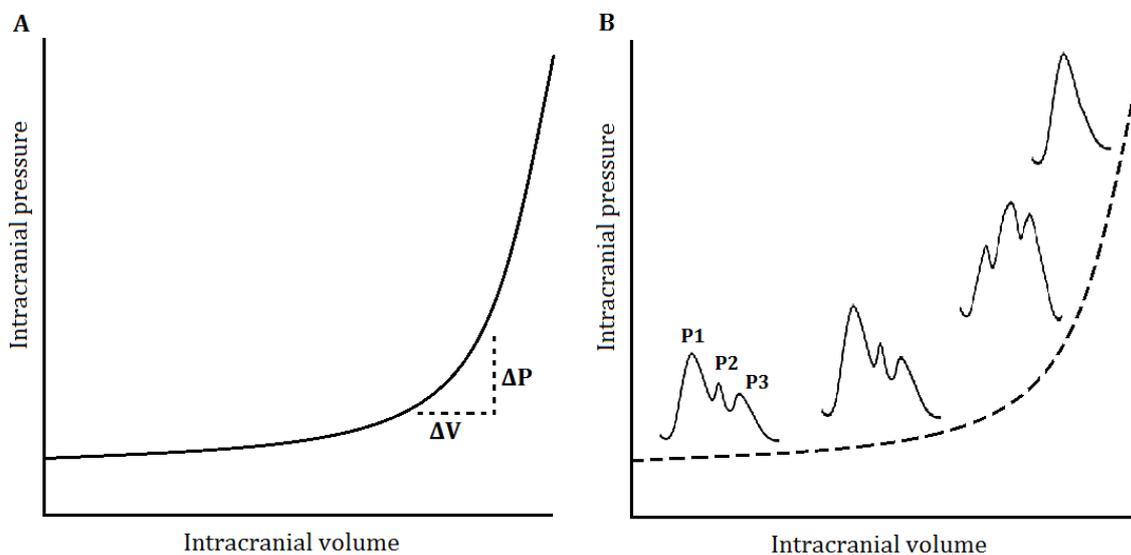


Figure 1: **A.** Intracranial pressure over intracranial volume. The compliance (dV/dP) is high as the intracranial volume is low, the pressure rise is small as the intracranial volume increases. There is no compliance as the intracranial volume is high. The beginning of the curve represents the stage in which the pressure-buffering capacity of the brain is functional [7, 26]. **B.** The hypothesized relation between the ICP waveform morphology related to an increase in the intracranial hypertension. The morphology of ICP-wave might elucidate information of the risk of development of intracranial hypertension linked to the intracranial compliance and pressure-buffering capacity of the brain [27].

The morphological changes of the waveform are the scope of this research. We hypothesize that a change in the morphology of the ICP wave precedes an increase in pressure. As the morphology of the ICP waveform is different under diverse (patho)physiologic conditions. Changes in the ICP waveform morphology are linked to the development of intracranial hypertension, acute changes in the cerebral CO₂ level, a decrease of the cerebral blood flow and changes in the craniospinal compliance [28]. The physiological waveform contains three peaks, as can be observed in figure 1B. The peaks are referred as P1 (percussion wave), P2 (tidal wave) and P3 (dicrotic wave). The origin of the ICP waveform is a complex interaction between the arterial inflow, venous return and the intracranial content. The waveform is primarily arterial with retrograde venous pulsation contributing to the latter components [29]. The exact origin of the morphology remains to be elucidated but some studies specify P1 as arterial pulsation, P2 as the intracranial compliance and P3 as aortic valve closure [28,30]. To the best of our knowledge there is no consistency in literature in the definition of P1, P2 or P3 peaks during pathological situations when peaks may merge.

In the first stage of increasing intracranial pressure the cranium encloses a compensatory mechanism for this increasing pressure, which can be explained by the Monro-Kellie doctrine [3]. An increase in intracranial volume, is compensated by a decrease of the venous volume or cerebrospinal fluid volume, preventing an increase of ICP. This pressure-buffering capacity of the venous blood and cerebrospinal fluid (CSF) is limited. The adaptive potential of this pressure-buffering capacity is represented in the compliance. The compliance is the change in volume over a change in pressure (dV/dP). In the first stage of a growing intracranial mass when the compliance is high, the rise in pressure is low as the volume increases. Beyond the pressure-buffering capacity there is a poor compliance of the brain: a small increase in volume induces a large increase in the ICP, figure 1A. An elevated ICP may impede cerebral blood flow (CBF) and cause ischemia [8]. The cerebral autoregulation regulates and maintains the CBF across a range of blood pressures. The autoregulation ensures that as the cerebral perfusion pressure (CPP) or mean arterial pressure (MAP) increase, the resistance of the cerebral vasculature increase by vasoconstriction and the CBF is maintained. Conversely, the CBF is maintained constant as the MAP or CPP decrease by vasodilation of the cerebral vasculature [1,31]. This mechanism of autoregulation functions only in a certain range, in this range a change in the systemic blood pressure or the cerebral perfusion pressure does not change the CBF [32]. Variations in the ICP-morphology are linked to a loss of intracranial compliance and dysregulation of the autoregulation mechanism [30,33].

We hypothesize that the waveform morphology of intracranial pressure pulses holds essential information about intracranial pathophysiology. The ICP-wave morphology can be regarded to a certain degree to reflect cerebral compliance. Changes in waveform characteristics, as the amplitude of ICP or rounding of pulse waveform can be detected in the signal prior to IH, figure 1B visualizes this hypothesis [34]. Automatic analysis of the ICP waveform may help to detect changes in the ICP waveform morphology and predict intracranial hypertension [28,35]. As current clinical practice is reactive and therapeutic interventions occur after a prolonged episode of elevated ICP, there is room for improvement of ICP monitoring systems that would provide more actionable information[35]. The goal of this research is to analyze the waveform morphology and develop an algorithm to automatically detect changes in the ICP waveform morphology and predict intracranial hypertension. By a machine learning approach, we hope to contribute to treatment of intracranial hypertension in TBI patients in a proactive manner [25]. This research will focus on prediction of intracranial hypertension by two different machine learning approaches. The first approach will predict IH based on a set of MOCAIP features describing the morphology of the ICP curve. We will develop an algorithm to detect the ICP sub-peaks, the algorithm will be learned over a set of by experts annotated ICP waves. A classification tree will be used to identify if a wave belonged to intracranial hypertension, the period before hypertension or a control segment. The second approach will elaborate on the use of neural networks to classify waves belonging to intracranial hypertension from normal ICP waves.

Method

Dataset

We performed a retrospective study in severe traumatic brain injury patients admitted to the ICU at the Radboud Universiel Medical Center, Nijmegen (RadboudUMC). Inclusion criteria were patients with severe TBI, continuously monitored by an intraparenchymal ICP probe. Patients were excluded as they obtained decompressive craniectomy.

Preprocessing

The raw ICP signal will be preprocessed to reduce the effect of noise. The first step is detection of individual ICP pulses, this was done by the use of a moving average filter.

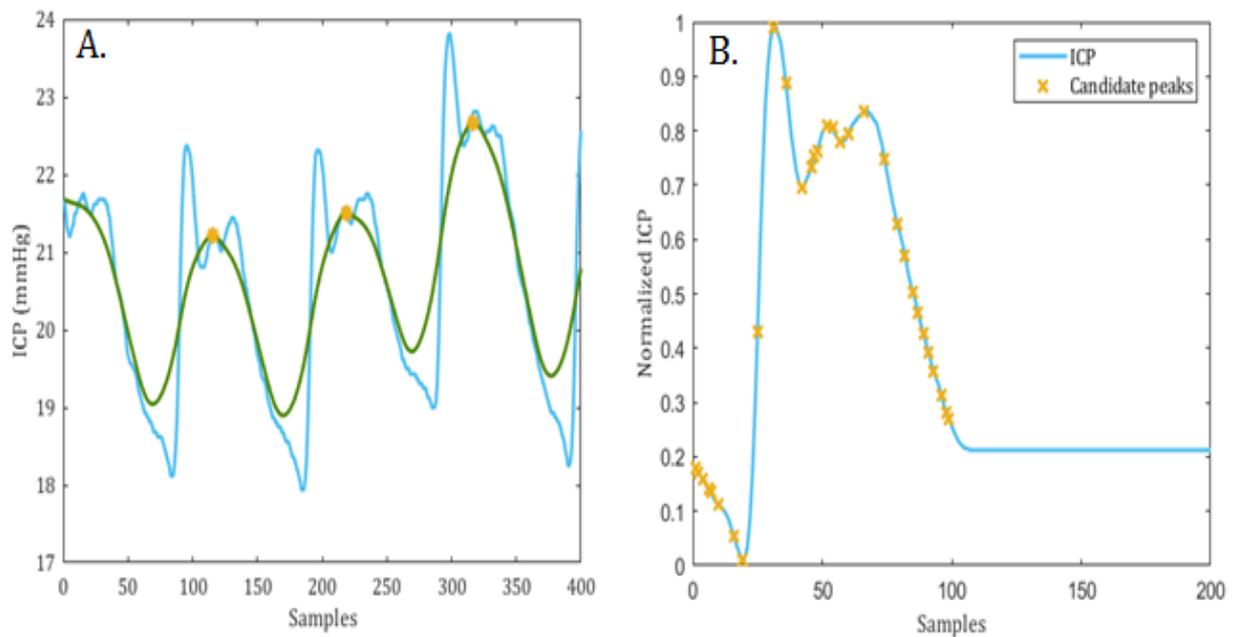


Figure 2: **A.** Raw ICP signal (light blue) with the low pass filtered ICP signal (green), the maximum of each period of the low pass filtered signal is marked (yellow). **B.** The averaged wave (blue) of the raw ICP signal in figure 2A. the candidate peaks were detected and marked by a yellow cross.

Moving average filter to detect heart beat

The ICP signal was filtered by a moving average filter using 30 samples and transfer coefficients $b = 0.0333$ and $a = 1$. This filter reduced the effect of high-frequency noise. The signal was filter forward and backward to eliminate phase shift. The low pass filtered signal is shaped like a sinusoid with one cycle per heartbeat [2]. The peaks of the low pass filtered signal were used to identify the location of one heartbeat. Individual ICP waves could be extracted as the location of

an ICP-wave was equal to the location of a heartbeat. Figure 2A visualizes the above described preprocessing steps for low pass filtering of the ICP signal.

Pulse preprocessing

The individual ICP waves should be represented as vectors with equal lengths. As the vector length of the ICP wave was >200 samples the extra samples at the end of the wave were discarded. If the wavelength was <200 samples the last value was repeated to fill the vector. A vector length of 200 was set as it was unpreferable to discard samples and lose information of the ICP wave. We assumed that a wavelength >200 samples was scarce as this corresponded to a bradycardia of <37 beats per minute. Figure 2B shows an example of an ICP wave filled with repeated values to 200 samples.

Pulse averaging and normalization

The individual ICP waves were often contaminated by noise and artefacts. The contamination originated from high frequency noise of electronic devices or patient movement. Individual ICP-pulses originating from one-minute of raw-ICP signal were averaged to reduce the influence of noise. The averaged wave was calculated by the mean of all individual waves in the one-minute signal. The averaged waves were used for the further analysis.

Each averaged wave was normalized between 0 and 1

$$\text{Normalized wave} = \frac{\text{wave} - \min(\text{wave})}{\max(\text{wave})} . \quad (1)$$

The normalization was performed to prevent a bias in classification between intracranial hypertension and no-intracranial hypertension by the absolute pressure of the averaged ICP wave.

Sub-peak detection as a regression problem

This section focuses on the problem of robust peak detection. Robust peak detection is important to extract morphological feature based on the position of the peaks from the ICP wave. These features will be used in this research to predict intracranial hypertension. The morphology of the ICP waves is extremely variable, this challenges accurate detection of the ICP peaks. A spectral regression algorithm will be implemented to deal with the variability in the ICP waves and assign the ICP sub-peaks (P1, P2, P3). The spectral regression algorithm requires a set a labelled ICP waves to train the algorithm. The labelled set will be constructed by experts of the RadboudUMC.

We evaluated the use of this labelled set as a gold standard by the intraclass correlation and the inconsistency in the labels assigned by the experts.

Expert labelling

We constructed a set of labelled averaged ICP-waves to train the spectral regression algorithm and test its performance. For construction of the test and training set experts in TBI care (intensivists, neurosurgeons and a technical physician) were asked to label sub-peaks in averaged ICP-waves by the use of a Matlab tool. They could assign 1, 2, or 3 peaks in each wave. As a peak was missing it obtained the label not a number (NaN). 730 waves were labelled by seven experts of the RadboudUMC.

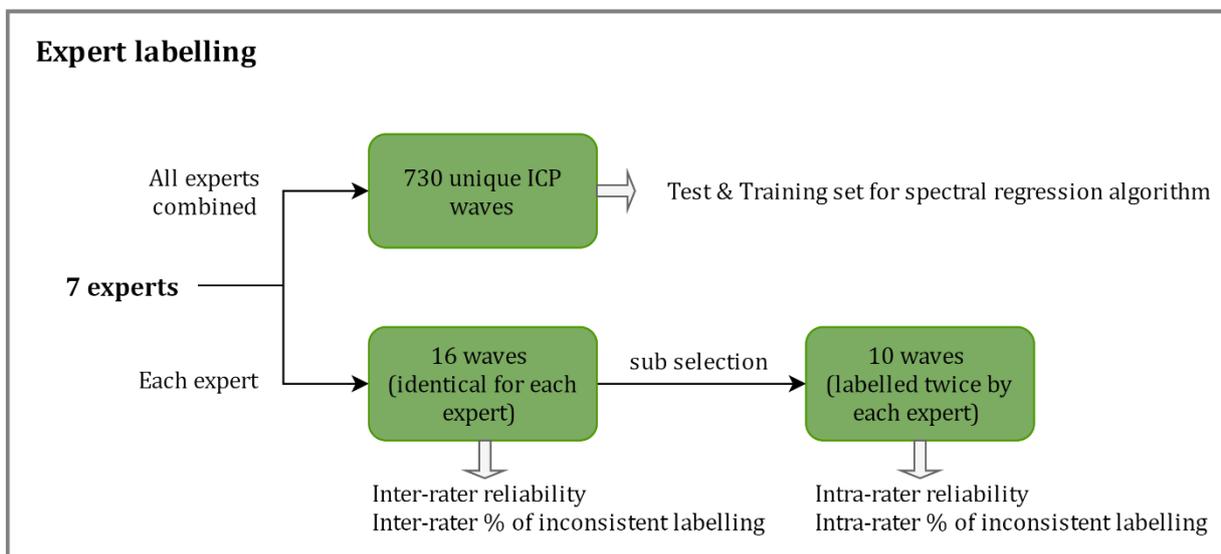


Figure 3: Overview of the number of dominant waves labelled by 7 experts of the Radboud UMC. The seven experts labelled together a total of 730 unique waves. These waves were used for testing and training of the spectral regression algorithm. All experts were asked to label 16 extra waves, these waves were identical for each expert. The experts labelled a sub-selection of these 16 waves twice. These waves were used to calculate the inter and intra-rater reliability and the percentage of inconsistent labelling.

Intraclass Correlation and inconsistent labelling

To evaluate the reliability and the consistency of the peaks labelled by the experts we calculated the intraclass correlation (ICC). The ICC was computed as an estimate of the inter-rater reliability and the intra-rater reliability. The interrater reliability reflects the variation between 2 or more raters who label the same set of peaks [36]. The ICC estimates and their 95% confidence intervals were calculated using SPSS (version 24, SPSS Inc, Chicago IL) based on mean rating (k=7), absolute agreement, 2-way random effects model [37]. The intra-rater reliability reflects the variation in the labelled peaks labelled by 1 rater across 2 or more trials. The intra-rater reliability was

calculated based on absolute agreement, One-way random effect model with a 95% confidence interval.

Inconsistent labeling was defined as the percentage of conflicting sub-peaks. We counted for the number of waves with inconsistency in the existence of a peak between the raters and calculated the percentage of disagreement (inter-rater % of inconsistent labelling). The P1 inconsistent labelling parameter was the percentage of waves where P1 was assigned by 1 or more experts and missed by others. The intra-rater percentage of inconsistent labelling was the percentage of conflicting subpeaks between the first and second time a rater labeled an identical wave. An overview of the number of waves labelled by the experts and reviewed for the inter-and-intra rater reliability and the inter-and-intra rater percentage of inconsistent labelling is visualized in figure 3.

Spectral regression

The spectral regression analysis applied in our research is a recent method, proposed by Scalzo et al. to detect ICP sub-peaks [38, 39]. The key idea of this method is to use a regression model combined with spectral graph analysis during sub-peak designation. The regression model is able to predict the most likely position of the three peak, $y = (P1, P2, P3)$ [38]. The spectral regression approach will be extended by a kernel to capture nonlinear relationships between the input and the output [40]. Using a supervised machine learning approach, we aim to find predictions for the sub-peak positions. Spectral methods are a powerful tool for dimensionality reduction. Dimensionality reduction has been a problem in machine learning as supervised machine learning algorithms degrade in performance when faced with many features that are not necessary for predicting the desired output [41].

Spectral regression is a method to solve discriminant analysis as a regularized regression problem

$$\alpha = \operatorname{argmin}_{\alpha} \sum_{i=1}^n (\alpha^T x_i - y_i) + \delta \|\alpha\|^2. \quad (3)$$

An elaborate description of the spectral regression analysis can be found in appendix 1. The regularization problem can be formulated as follows

$$\alpha = (XX^T + \delta I)^{-1} X^T y. \quad (4)$$

Where X is a matrix formed by the input waves $[x_1 \dots x_n]$, α is the eigenvector, I is the identity matrix and δ is the regularization parameters, we used $\delta = 0.001$. This formula can be solved using Cholesky decomposition [21].

We used Kernel spectral regression (KSR) to extend this method to nonlinear problems. This allows to use a linear regression analysis to solve a nonlinear problem. The kernel maps the input waves x to a higher dimensional space. We used a Gaussian kernel (K):

$$K_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right). \quad (5)$$

Where σ is the standard deviation of the kernel. Similar to the spectral regression approach we use Cholesky decomposition to obtain the vector α ,

$$r = \text{chol}(K + \delta I) \quad (6)$$

$$\alpha = r \setminus (r^T \setminus y). \quad (7)$$

Now we can find the embedding function:

$$\text{Embedding function} = K * \alpha \quad (8)$$

where α is the eigen vector that minimized the residual sum of square error in equation 3.

We calculated the class centers (mean of the embedding functions) of the wave with the same label in the training set [42].

The Kernel of the test set was constructed, in a similar way as the Kernel from the training set, equation 5. The embedding function of the test set was calculated by equation 8. The corresponding labels were assigned by minimization of the Euclidian distance between the embedding functions of the test set and the class centers of the embedding functions of the training set. The labels of the class center that minimized the Euclidian distance were assigned to the waves in the test set. The KSR algorithm assigned 1, 2, or 3 ICP-subpeaks to the averaged ICP waves.

Train algorithm

The KSR algorithm was trained by, the leave one out crossvalidation method. During the leave one out training the training set contained all averaged waves except for the averaged waves belonging to one patient. The test set contained the averaged waves of the last patient. The cross-validation was performed 31-times.

Detect candidate peaks

We will search for candidate sub-peaks in the averaged waves. Each candidate peak was susceptible to become one of the three ICP sub-peaks. The first and second derivative of the averaged waves were calculated to identify each candidate peak. A sample was located as candidate peak as it met the following definitions. The first derivative of the signal is 0. The second derivative changes from concave to convex or vice versa. The second derivative was convex as its sign was >0 , the second derivative was concave as its sign was <0 . An example wave can be observed in figure 2B.

Sub-peaks

The ICP sub-peaks were assigned based on the minimal distance between the candidate peaks and the peaks predicted by the KSR algorithm. The candidate peak closest to the first regression peak was assigned as P1. This process was repeated for P2 and P3, if assigned by the spectral regression algorithm.

Position of sub-Peaks

The experts and the spectral regression algorithm could assign 1, 2 or 3 sub-peaks to a averaged wave. We renamed the sub-peaks to compare the peaks labelled by the expert to the corresponding peaks of the spectral regression algorithm in case they assigned a different number of sub-peaks. When one peak was assigned to a wave we called this peak P2. In case two peaks were assigned, the position and the amplitude of sub-peak determined if we called the peaks P1 and P2 or P2 and P3.

Prediction accuracy

Sensitivity, specificity, accuracy and the prediction error were used to evaluate the performance of the spectral regression algorithm.

The prediction error was calculated for the true positive peaks. The prediction error in milliseconds (ms) is the error between the actual position of the peaks, assigned by experts, y and the predicted peak position \hat{y} by the spectral regression algorithm, equation 9 [39]. The average prediction error is the mean of $e1, e2, e3$

$$\{e1, e2, e3\} = \frac{1}{n} \sum_{i=1}^n |\hat{y} - y|. \quad (9)$$

MOCAIP metrics

We used Morphological clustering and analysis of ICP (MOCAIP) to extract features from the ICP-wave. MOCAIP analyzes morphological changes in the ICP-pulse and provides intuitive parameters of the ICP-pulse to the clinician. 27 metrics were calculated allowing characterization of the ICP pulse morphology by the pulse amplitude, time intervals among subpeaks, curvature, slope and decay of time constants [18]. Figure 4 visualizes the MOCAIP metrics in an ICP -wave. metrics were calculated depended on the peaks assigned by KSR algorithm. Metrics were assigned empty as they were based on a non-existent subpeak. We added two extra metrics to the original MOCAIP metrics, the maximal amplitude and the maximal slope. These metrics were added as they did not include the position of P1, P2 or P3. They are robust for limitations of the spectral regression algorithm. An extensive description of the 27 metrics is described in appendix 2.

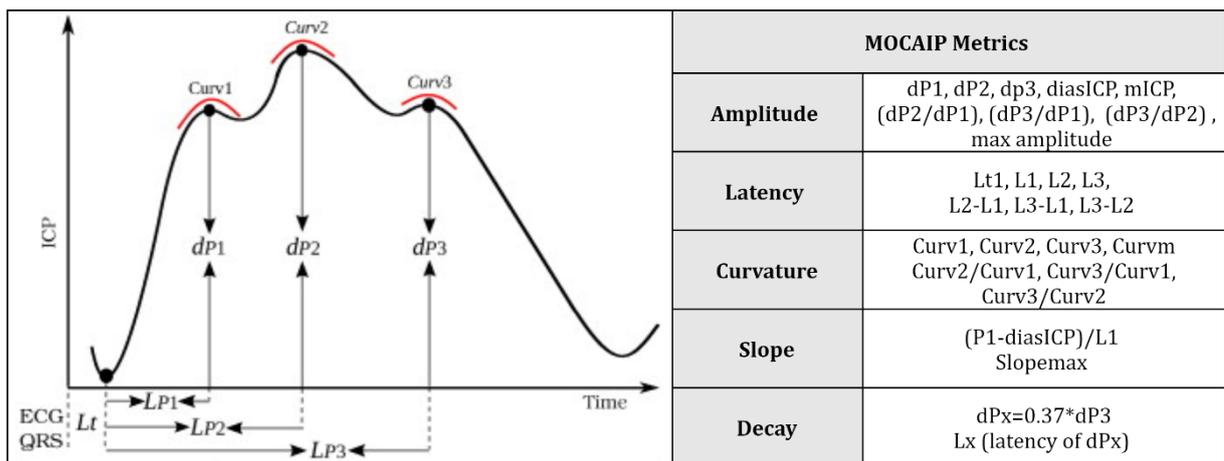


Figure 4: Visualization of the 25 original MOCAIP metrics from an averaged ICP wave. The metrics were based on the amplitude, latency, curvature, slope and decay [35].

Segment selection

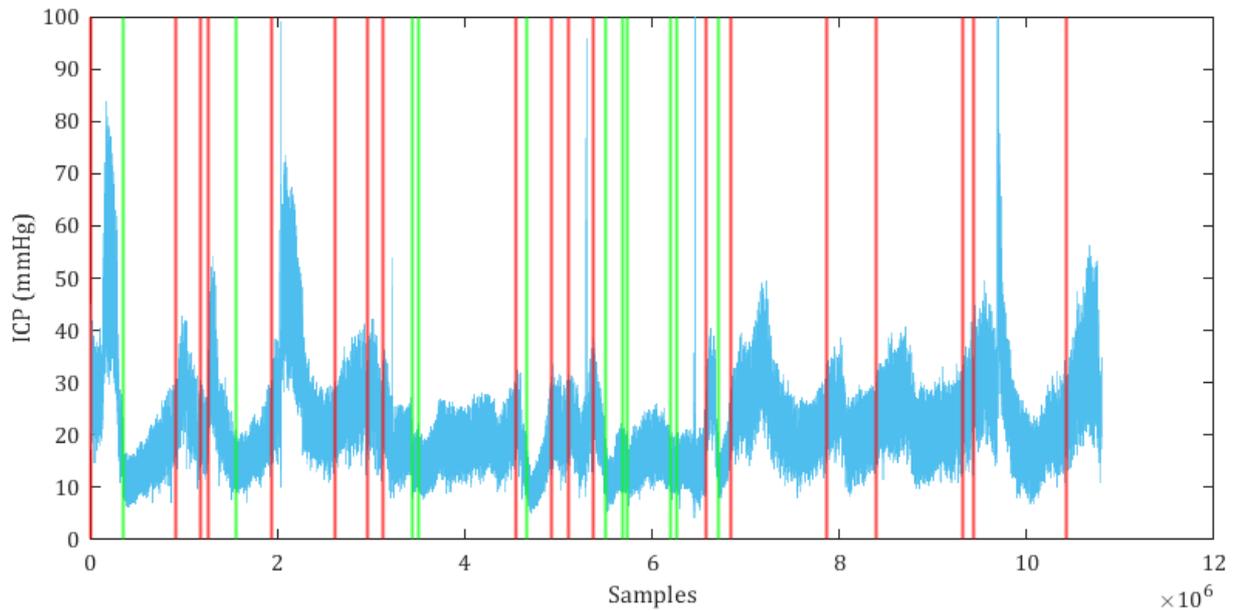


Figure 5: Selection of intracranial hypertension segments and no-intracranial hypertension segments. The red vertical lines mark intracranial hypertension periods, ICP > 22 mmHg for > 5 min. The green vertical lines mark periods without intracranial hypertension, the ICP was < 15 mmHg for a minimum of 5 minutes.

To predict intracranial hypertension, we had to select waves belonging to IH, no-IH and waves leading to IH. To obtain these waves two types of segments were selected from the raw ICP signal. Segments to classify intracranial hypertension versus no-intracranial hypertension (figure 5) and segments to predict intracranial hypertension (figure 6). All segments were selected using a sliding window, which averaged the mean pressure over 1 minute, shifted to the next window and averaged the mean pressure again. The segments in figure 5 to classify IH and no-IH waves had to fulfill the following requirements. IH segments: the ICP should be > 22 mmHg for > 5 consecutive minutes. There should be minimum period > 5 min between the IH segments. No-IH segments: the ICP < 15 mmHg for > 5 min, >5 min between no-IH segments. A label was assigned to the segments for classification. The label 1 was assigned to IH segments and 0 to no-IH segments. 181 segments of intracranial hypertension and 282 no-IH segments were selected.

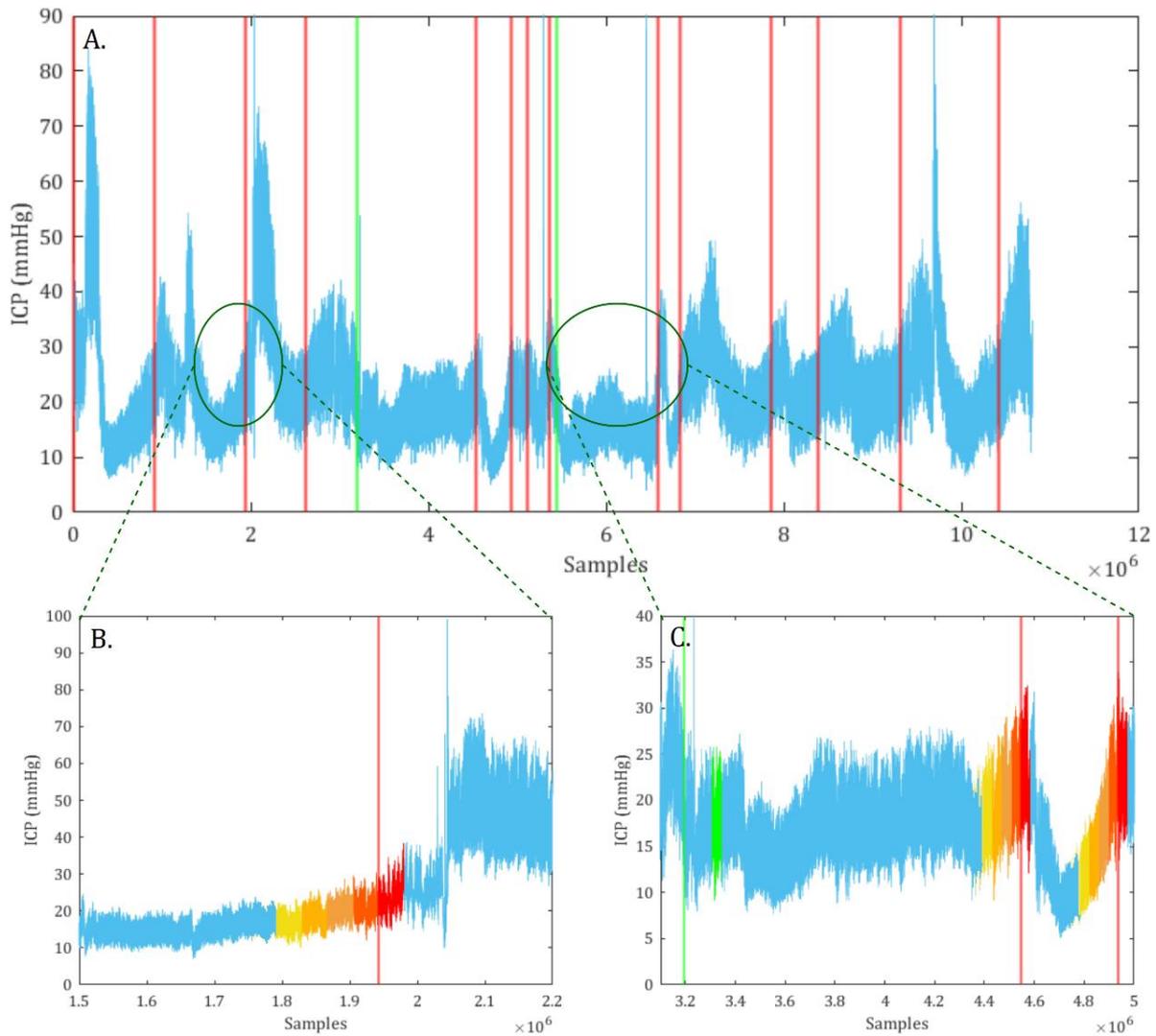


Figure 6: ICP signal over time (samples), intracranial hypertension and control segments were detected by a sliding window. The start of an IH segments is indicated by a red vertical line. The start of the control segments is indicated by a green horizontal line. We zoomed in on two parts of the signal. **B.** A segment leading the hypertension. 5 minutes of hypertension are showed in red, the pre-IH segments are visualized by orange/yellow colors. Each color indicates a pre-IH segments of 5 minutes. **C.** Zoomed in on a control segment. The control segments were selected 15 minutes after the start of a period with a minimum of 90 minutes ICP <22 mmHg. This control period was followed by two segments leading to hypertension.

Segments leading hypertension had to meet refined requirements. Figure 6 illustrates the requirement for segment selection. The ICP should be > 22 mmHg for a minimum of 5 consecutive minutes, this is the start of IH, marked by the red vertical lines. Label 1 was assigned to the IH segments. The ICP should be < 22 mmHg in the period before IH. The period before IH was split in 4 segments of 5 minutes. These segments obtained the following labels, 0-5 minutes pre-IH (Label 2), 5-10 minutes (Label 3), 10-15 minutes (Label 4) and 15-20 minutes (Label 5) prior to IH. These pre-IH segments are marked in figure 6B and C by orange and yellow colors. Control

segments were selected in the same signal. The control segments were selected from a control period where the ICP was < 22 mmHg for > 90 min. The 5- minute control segments were selected 15 minutes after the start of the 90minute segment. The start of the control periods was marked by green vertical lines. The control segment 15 minutes after the start of the control period was marked green as visualized in figure 6C. Multiple pre-IH/IH segments and control segments could be selected from the same patient. In total 96 pre-IH/IH segments and 79 control segments met the criteria. Only segments from patients having both a pre-IH/IH segment and control segment were included in further analysis. This resulted in a selection 85 pre-IH/IH segments from 19 patients. These patients had a median of 4 [1.25 5.75] IH episodes.

Classification tree - Morphological features

Statistical analysis of MOCAIP metrics

The MOCAIP metrics were calculated for all averaged waves in the selected segments. The individual MOCAIP metrics were checked for gaussian distribution. As the data was not normal distributed, a Kruskal Wallis test was performed. We tested for significance between the MOCAIP metrics in segments leading to IH. The changes were reported for IH, Pre-IH (0-5, 5-10, 10-15, 15-20 minutes before IH) and a control segment. The control segments were selected from the same patients. The null hypothesis was rejected as $P > 0.05$. 85 segments leading to IH were included in this analysis. This analysis was performed in GraphPad Prism.

Classification tree

Classification was performed to distinguish IH waves from no-IH waves and to classify the pre-IH segments based on the MOCAIP metrics. A classification tree package in Matlab was used for this analysis. The classification tree classified the waves into classes 0 or 1, no-IH or IH. For classification of the pre-IH and control segments the classification tree classified in 0, 1, 2, 3, 4, 5, 6. The parameters in this classification model were tuned by the minimal leave size and the number of function evaluations to obtain the best classification results. Different values for the min leaf size were tested to find a balance between the most optimal solution for the training and test set, as we should prevent overfitting on the training set [43]. A leave one out method was used for training and classification of the IH/pre-IH and control waves. The classification performance was evaluated by the sensitivity, specificity, accuracy, negative predictive value (NPV) positive predictive value (PPV), false positive rate (FPR) and the false negative rate (NPR).

Neural Network – Averaged wave

The use of a Neural Network was explored for classification of IH and no-IH waves based on their morphology. Inputs of the previous MOCAIP classification experiment were metrics describing

morphological changes of the wave. This neural network experiment uses the averaged, normalized wave as its input. It is not limited to pre-defined features for classification, the neural network had the freedom to recognize any properties that are characteristic for IH [25]. Two types of neural network were evaluated for their classification performance, we used a Long Short Term Memory (LSTM) network and a Convolution neural network (CNN). The LSTM network relies on temporal changes of the signal. As the input were averaged ICP waves we investigated the use of convolution neural network to discard this problem. The CNN network was constructed using 12 layers: 3 convolutional layers, 2 batch normalization layers, 2 relu layers, 2 dropout layers, 1 fully connected layer, 1 softmax layer and 1 classification layer. The network was trained over 50 epochs with a mini-batch size of 15 and 10 hidden units.

The LSTM network was constructed using to 2 LSTM layers (one sequence and one last), 1 sequence input layer, 2 fully connected layers, 1 softmax layer and 1 classification layer. The network was trained over 5 epochs, with a mini-batch size of 15 and 10 hidden units.

Results

Dataset

The dataset used in our research originated from 53 traumatic brain injury patients admitted to the intensive care department of the Radboud-UMC, Nijmegen. These patients were admitted to the ICU between Mai 2017 and August 2019. 22 patients were excluded from the analysis as they obtained decompressive craniectomy.

Table 1: Patient characteristics of TBI patients at the Radboud UMC, Nijmegen.

Descriptive		
TBI patients (with ICP monitoring between Mai 2017 – Aug 2019)		53
Sex	Male	39
	Female	14
Age in years	(median (IQR))	44 (30 - 58)
Decompressive craniectomy		22

Sub-peak detection as a regression problem

Expert Labelling - Inter-intraclass correlation and inconsistent labelling

The performance of the experts to label the ICP sub-peaks P1, P2, P3 was evaluated. Table 2 presents the intra class correlation and the percentage of inconsistent labelling of the three ICP-subpeaks by seven experts. The intraclass correlations are close to one for all subpeaks. What stands out in table 2 is the high inconsistent labelling for P1, as well as the inconsistency of 0% for P2.

Table 2: The intra- class correlation of the sub-peak positions P1, P2, P3 labelled by seven experts. The intra-class inconsistent labelling is the percentage of identical peaks labelled by one or more experts and missed by others. The inter-class percentage of inconstant labelling is the percentage of inconsistent assigned peaks between the first and second time a rater labelled a wave.

Peak	Intra-class correlation (95 % CI)	Intra-class, Inconsistent labelling (%)	Inter-class, Inconsistent labelling (%) (ME [IQR])
P1	0.999 (0.997 – 1)	43.75%	10 % [2.5% 10%]
P2	0.990 (0.981 – 0.996)	0%	0% [0% 0%]
P3	0.997 (0.992 – 0.999)	18.75%	10% [0% 10%]

A complete overview of the results for the intraclass correlations and the percentage of inconsistent labelling of the three ICP subpeaks for the individual experts is reported in appendix 2. It is apparent from these results that only one expert was consistent in peak labelling for all waves. As can be observed in table 2, P1 and P3 had again the highest inconsistency, the median of inconsistent labelled P1 and P3 peaks was 10% for the experts between the first and second time they labelled a similar wave. The interclass correlations for all raters and peaks are close to 1. This indicates that the experts labelled precisely.

Spectral regression

Peak prediction

We calculated the error between the peaks labelled by the experts and assigned peaks in the 31 leave one out cross validation method of the KSR algorithm. The peak prediction error was 2.3 samples (18 ms) for P1, 5.8 samples (46 ms) for P2 and 6.1 samples (48 ms) for P3. The average prediction error over the three sub-peaks was 4.7 samples (38 ms). As table 3 shows, the sensitivity, specificity and accuracy of P1 was quite low. The sensitivity for P2 was 100% and the accuracy 99%. The specificity was undetermined for P2 as there were no true negative or false positive peaks because P2 was assigned and present in all waves. The presence of P3 was more variable but better compared to P1. An interesting observation is that the largest confusion clearly exists in the assignment of P1, which is in consonance with the inconsistency in expert labelling.

Table 3: Sensitivity, specificity and accuracy of the sub-peaks P1, P2 and P3 predicted by the leave on out cross validation method. The peaks labelled by the spectral regression algorithm were compared with the peaks labelled by the experts. The ground truth were the labels assigned by the experts.

Peak	Sensitivity	Specificity	Accuracy
P1	65%	50%	57%
P2	100%	-	100%
P3	89%	23%	79%
P123	88%	43%	78%

A selection of ICP waves labelled by expert (marked in yellow) and the sub-peaks predicted by the spectral regression algorithm (marked in blue) are presented in figure 7. In figure 7A, C and D the expert and KSR marked the same number of sub-peaks. Remarkable is figure 7B where P1 is marked by the KSR but not assigned by the expert. There was a large deviation in the position of the sub-peaks in figure 7D. The peaks marked by the KSR did not correspond to the right position of the sub-peaks.

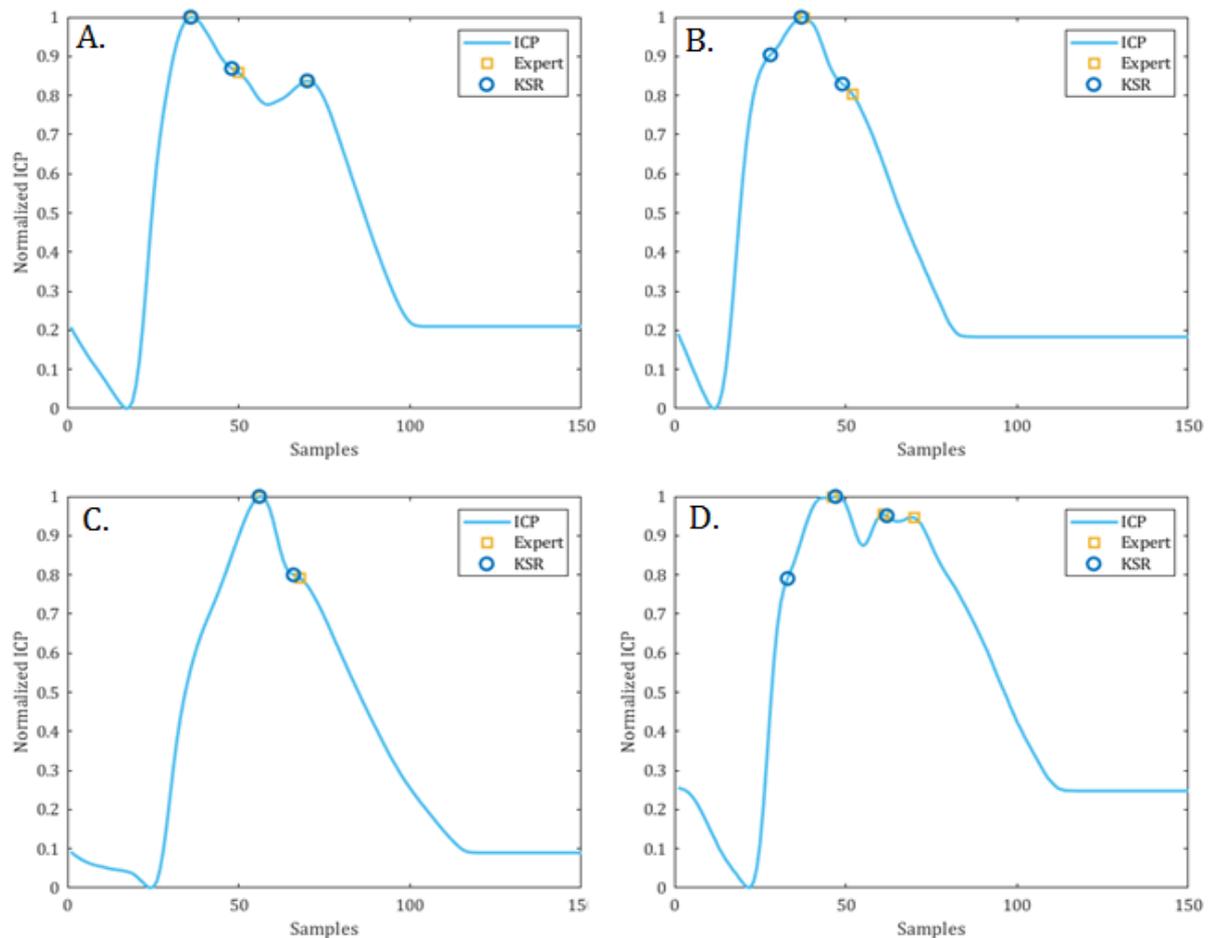


Figure 7: A selection of the results of the Kernel Spectral Regression algorithm (KSR). Dominant ICP waves with the manual labelled peaks by the expert in yellow and the labels predicted by the algorithm in blue.

The performance of the spectral regression algorithm improved, especially for exceptional ICP waves, as we randomly split the waves over the test and training set. We tested for 80% of the labelled waves in the training set and 20% in the test set (80-20 training). Figure 8 shows the results obtained by the leave one out training approach and the 80-20 training approach for an exceptional but physiological ICP wave. The leave one out approach assigned three unrealistic peaks, figure 8A. The 80-20 approach assigned 2 realistic peaks corresponding to the peaks assigned by the experts, 8B.

This illustrates a limitation of the current KSR algorithm. The prediction errors, sensitivity, specificity and accuracy of this 80-20 training approach can be found in appendix 3. The poor results of the leave one out training method can be explained as this type of wave occurred just in one patient in our dataset, the KSR was not trained on this type of wave. The KSR assigned the peaks of the closest class center. In the 80-20 approach the KSR was trained on a similar wave and performed accurate assignment of the sub-peaks. According to these results, we can infer that the algorithm worked but its generalizability is limited by the small amount of training data.

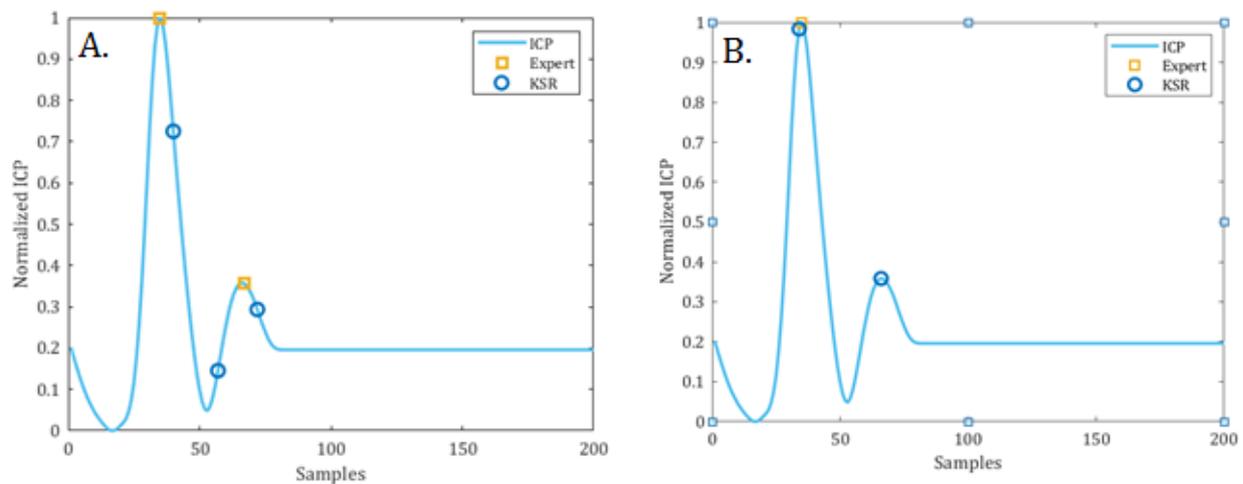


Figure 8: **A.** Averaged ICP wave where we marked the peaks assigned by the KSR algorithm and the experts. The peaks predicted by the KSR algorithm resulted from the leave-one-out training method. The peaks assigned KSR are aberrant. **B.** Averaged ICP wave where the peaks assigned by the KSR algorithm are the result of the 80-20 training approach. The peaks assigned by the KSR algorithm and the expert have identical positions.

MOCAIP metrics

Statistical analysis of MOCAIP metrics

A Kruskal Wallis test was conducted to test for significance of the 27 MOCAIP metrics in the IH, pre-IH and control segments, table 5 gives an overview of the results. 85 segments leading to IH and 85 control segments were included in this analysis. The colors in table 5 indicate statistical significance, green states statistical significance and yellow states no statistical significance. 20 metrics were significant different between the IH, pre-IH and control waves. These metrics changed over time in the period before IH. L1, L2, RatioL12, Curv1, Curv12, Curv31 and Lx were never significant. 11 to 18 metrics were significant between the IH vs 5, 10, 15, 20-minute segments before IH and control. The amount of significant metrics between the pre-IH segments before IH and control was limited. There were no significant metrics in 10 vs 15, 10 vs 20 and 15 vs 20 minutes. These results can be found in appendix 5. We can assume that morphological changes over time were limited in these pre-IH segments. Closer inspection of the table showed significance of the parameters mICP, diasICP, Ratiod13, Ratiod23 and Curv2 between the control segment and pre-IH segments. Some changes in the morphology might have occurred between the control period and the pre-IH periods. These results support the potential to distinguish waves leading to intracranial hypertension from control waves.

Table 5: Summary statistics of the Kruskal Wallis test. The green color indicated statistical significance between the two groups for the corresponding metric. Yellow indicates no statistical significance between the two groups. The first column is the P-value, a P-value <0.05 states a significant change over time for the corresponding metric in the period from control to intracranial hypertension.

	P-value	IH vs 5	IH vs Control	5 vs Control	10 vs Control	15 vs Control	20 vs Control
mICP	<0.0001						
diasICP	<0.0001						
dP1	<0.0001						
dP2	<0.0001						
dP3	<0.0001						
Ratio12	0.0028						
Ratio13	< 0,0001						
Ratio23	< 0,0001						
Amplitude	< 0,0001						
Lt	0.0007						
L1	0.9282						
L2	0.8573						
L3	0.0005						
RatioL12	0.6063						
RatioL13	0.0014						
RatioL23	0.0007						
Curv1	0.9169						
Curv2	< 0,0001						
Curv3	< 0,0001						
Curvm	< 0,0001						
Curv12	0.1956						
Curv31	0.3033						
Curv32	0.0004						
Slope	< 0,0001						
Lx	0.1362						
dPx	< 0,0001						
Slopemax	< 0,0001						

Classification tree –Morphological features

A supervised learning approach was used for classification between IH and no-IH waves based on the MOCAIP metrics. The aim of this section was to determine the feasibility to distinguish IH-waves and no-IH waves based on MOCAIP metrics describing the morphology of the curve. The metrics mean ICP and diastolic ICP were excluded from this analysis.

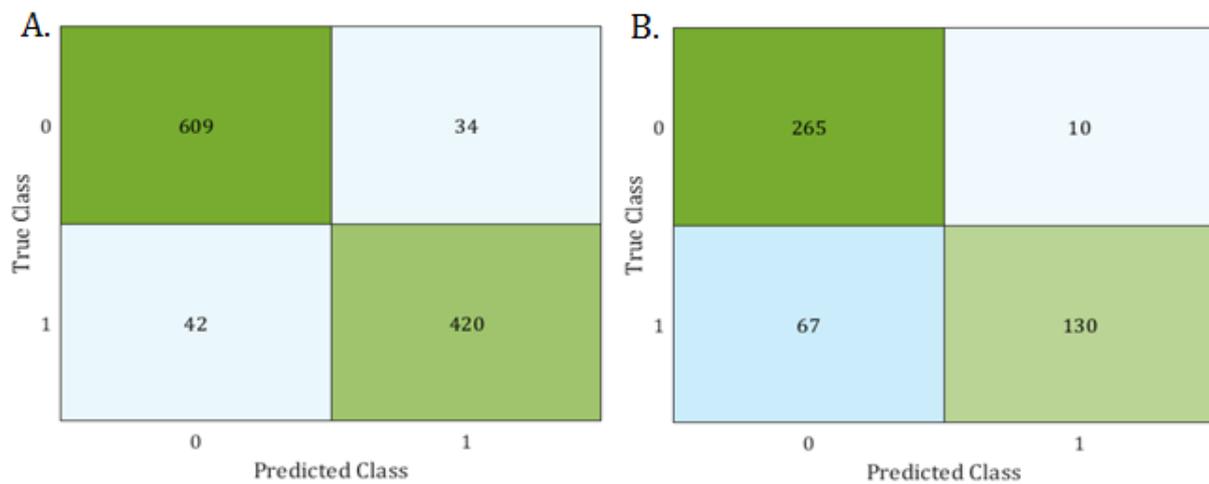


Figure 9: **A.** Confusion matrix of the training set for classification of IH (Class 1), and no-IH (Class 0) based on the MOCAIP metrics. **B.** Confusion matrix of the test set for classification of IH and no-IH based on the MOCAIP metrics.

The results of the classification tree for classification of IH-waves and no-IH waves based on the morphology of the curve can be observed in the confusion matrices in figure 9. The elements in the diagonal (in green) are correctly classified. The elements out of the diagonal are misclassified. A minimal leaf size of 18 was used as it had the best performance for both the test and training set. IH waves were classified as 1, no-IH wave were classified 0. The accuracy for the test set was 93% and the accuracy for the training set was 84%. The sensitivity for correct classification of waves belonging to IH and no-IH was respectively 91% for training and 66% for testing. The specificity was 95% for training and 96% for testing.

Figure 10 visualizes the predictive power of the MOCAIP metrics for classification of IH and no-IH waves. The predictive importance was defined as the change in the sum of the mean squared error due to a metric, divided by the sum of the number of branch nodes. The slope had the highest predictive value. dP3 was the second most important and Lt and dPx the third most important metrics.

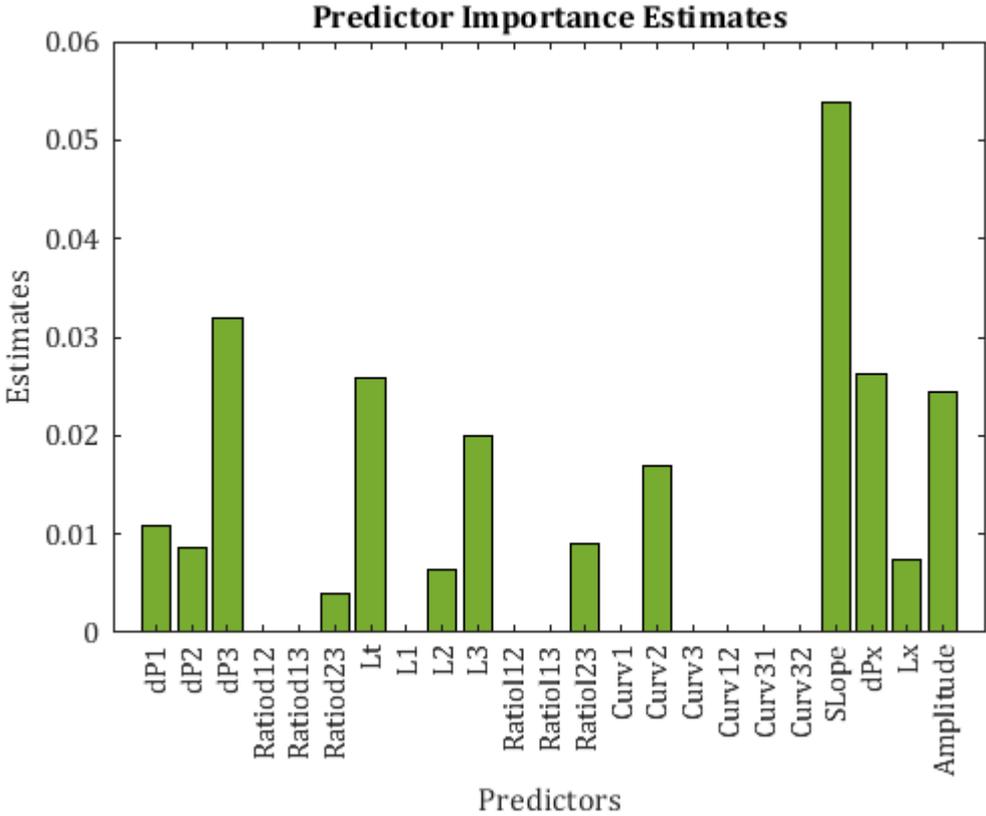


Figure 10: Predictor importance estimates of the MOCAIP metrics for classification between IH (ICP > 22 mmHg) and no-IH (ICP <15 mmHg).

The most challenging classification is to classify the IH, pre-IH and control waves based on the MOCAIP metrics. As pre-IH waves and control waves could be classified correctly, prediction of an episode of intracranial hypertension may be feasible. The 20 significant metrics in table 5 were included in this classification. The confusion matrix in figure 11A, shows the leave one out classification results for the test set. Class 1 represents IH, 2 represents 0-5 minutes before IH, 3: 5-10 minutes before IH, 4: 10-15 minutes before IH, 5: 15-20 minutes before IH, 6: Control waves. The overall accuracy was 42%. Figure 12, presents an overview of the accuracies (acc), sensitivities (sens), specificity (spec), positive predictive value (PPV), negative predictive value (NPV), true positive rate (TPR) and the false positive rate (FPR) of the different classes.

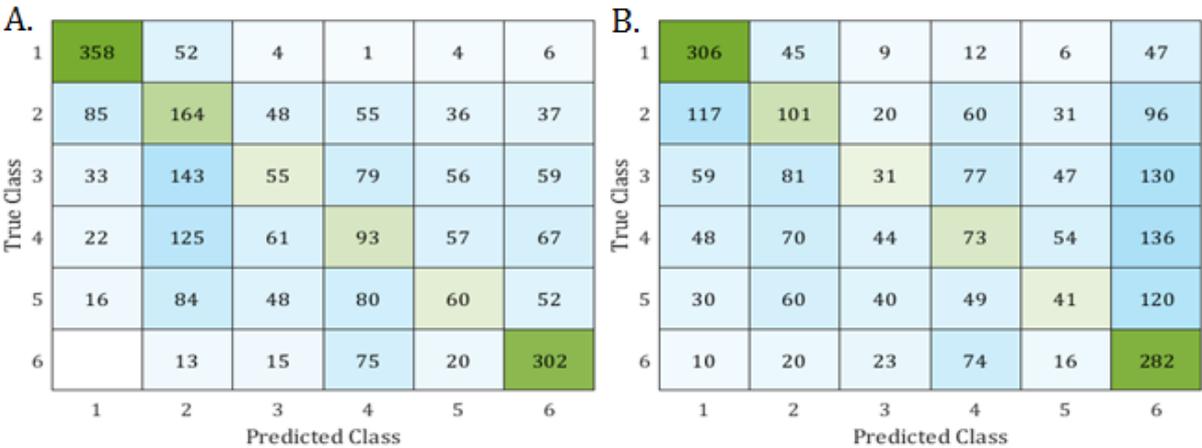


Figure 11: **A.** Confusion metrics of the test set, prediction of the classes 1: IH, 2: 0-5 min before IH 3: 5-10 minutes before IH, 4: 10-15 minutes before IH, 5: 15-20 minutes before IH, 6: Control segments. Classification was performed including all significant MOCAIP metrics. **B.** Confusion metrics of the test set of the same classes as in figure A. Classification was performed with all significant metric based on the morphology of the curve (excluded mICP and diasICP).

The obtained results for classification between waves leading to hypertension (pre-IH and IH) vs control waves are promising as the obtained a sensitivity of 89%, specificity of 71% and accuracy of 86%. The classification of the pre-IH segments at their specific timing is limited as the obtained sensitivities were between 13 and 39%. These classifications included the mICP and diastolic ICP. The mICP and diasICP were the most important predictors for classification. We performed the same analysis while excluding these 2 metrics. The confusion matrix based on metrics describing the morphology of the curve is visualized in figure 11B. What stands out in this matrix is the increase in the amount of false negative waves. They were predicted as a control wave while they often belonged to the pre-IH classes. The overall accuracy decreased to 34%. The sensitivity, specificity and accuracy for classification of the IH and pre-IH vs the control segments were respectively 74%, 66% and 73%. dP1 was the most important predictor.

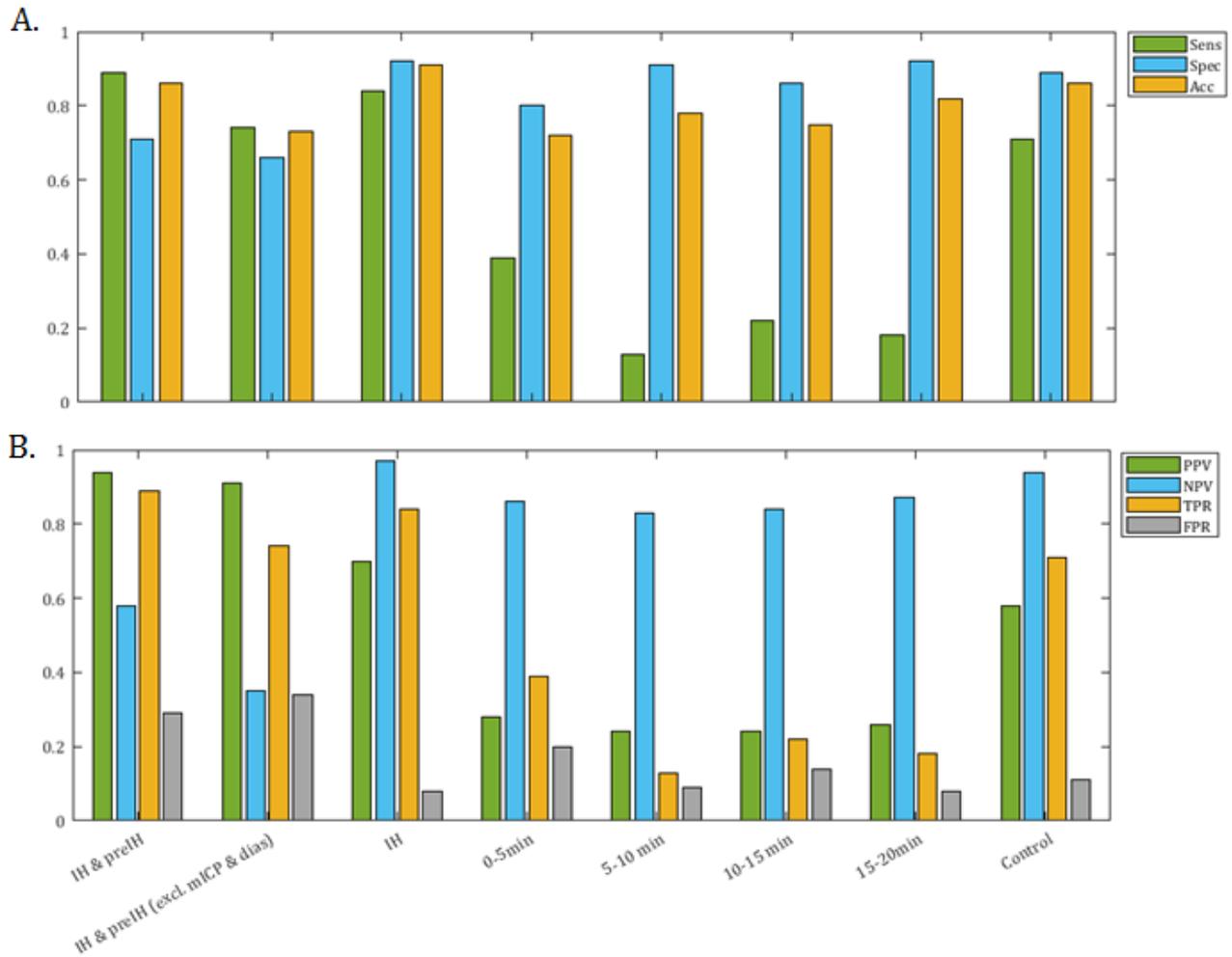


Figure 12: *A.* Sensitivity, Specificity and Accuracy of waves leading to hypertension (IH + pre-IH), waves leading to hypertension without mICP and diasICP, IH waves, pre-IH waves at their specific timing and classification of the control waves. The results of the classification tree based on the MOCAIP metrics. *B.* The positive predictive (PPV), negative predictive value (NPV), true positive rate (TPR) and false positive rate (FPR) of the same classes and waves as in figure A.

Neural Network – Averaged wave

The classification results for the LSTM neural network were inconsistent. We trained the LSTM network by averaged waves with the label 1 IH and 0 no-IH. The classification results can be observed in table 6. The results of the convolution neural network were sensitive for overfitting on the training set. A dropout layer was added to prevent overfitting.

Table 6: Classification results (Sensitivity, Specificity and Accuracy) of intracranial pressure waves and no – intracranial pressure wave by a LSTM neural network and a CNN network.

	Test - LSTM	Training - LSTM	Test - CNN	Training- CNN
Accuracy	70%	69%	58%	77 %
Sensitivity	97%	58%	47%	85%
Specificity	66%	88%	60%	75%

The results for the test and training set are very disparate for both the LSTM and CNN. As the sensitivity from the LSTM test set was very high, we could assume that as the network classified a wave as IH it should be true. Closer inspection rejects this assumption as the sensitivity for training was limited to 58%. The results of the LSTM network are slightly better compared the CNN network.

Discussion

Our goal was to predict intracranial hypertension. We developed an algorithm to accurately detect and extract morphological information from the ICP wave. We tried to detect the ICP sub-peak positions P1, P2 and P3 by the utilization of a machine learning approach called spectral regression analysis. The spectral regression method was a very efficient method to extract the sub-peaks positions after training with a training set labelled by experts. We evaluated how experts assigned ICP subpeaks to an ICP waves. Interesting insights were obtained as the experts were inconsistent in what they called a sub-peak. 27 (MOCAIP) metrics describing the morphology of an ICP wave were extracted from each averaged wave and used for classification between IH, pre-IH and control waves. Our promising results suggest the potential to distinguish waves leading to intracranial hypertension from control waves.

Preprocessing

The performed preprocessing steps focused on selection of individual ICP peaks to create a robust averaged peak. Artefacts or segments with noise were not removed before analysis. Instead, an averaged wave was constructed as the mean of one minute of raw individual waves. It is recommended to improve this preprocessing step in future to prevent forming of non-physiological signals. A potential solution is to cluster waves and select the largest cluster as the averaged wave. Additionally the waves could be compared with a reference library with valid ICP waves, as the averaged wave correlates with a reference ICP wave, the averaged wave could be identified as physiological and valid [39]. Improved preprocessing will result in less variation in the averaged waves, this will improve the performance of the spectral regression algorithm and increase the reliability of the MOCAIP metrics.

Expert labelling

An important finding was found in the peak labelling consistency. There was a large disagreement in the presence of ICP sub-peaks. In 43.75% of the waves the experts did not agree over the presence or absence of the P1 sub-peak. The agreement was better for P3, as the experts disagreed in 18.75% of the waves. The agreement was excellent for P2 where the experts agreed in 100% of the waves about the presence of P2. This finding was not unexpected because the label P2 was always assigned to one of the subpeaks. It is somewhat surprising that the expert opinion over the sub-peak existence is so disparate. There is no gold standard for sub-peak positioning in literature. This implies that we should be careful in the intensive care unit in the discussion about ICP and the position of the sub-peaks. Prior studies focused on accurate detection of the ICP sub-position P1, P2, P3 [10, 28, 29, 44]. Very little was found in the literature on the question what

should be called a sub-peak. Being aware of this we should review the use and naming of P1, P2, P3 in prior studies and carefully interpret their results. A potential solution to limit influence of the missing gold standard is to develop an unsupervised learning algorithm. This algorithm will be insensitive for the missing gold standard as no expert opinion is included in the peak assignment. A potential solution was proposed by Lee et al. They developed an algorithm based on peak clustering to assign peaks in individual pulse morphologies [45].

We asked experts to label ICP peaks in preprocessed waves visualized on a laptop screen. The displayed ICP-wave was large compared to an ICP-wave on the monitor at the ICU. This allowed the experts to distinguish more variations in the ICP wave than they would see on the monitor. The experts were not familiar with such a detailed image of an ICP-wave, causing confusion in the interpretation and assignment of the sub-peaks. In addition, as the clinician reviews the ICP-signal at the ICU unit the MAP signal is visualized in the same window. The MAP signal might be used to interpret the ICP signal. The dicrotic notch of the MAP should be aligned with the dicrotic notch of the ICP signal. This marker can be useful to determine the sub-peak positions. The dicrotic notch is associated with the closure of the aortic valve at the end of the ventricular systole. The wave components before the dicrotic notch are assumed to be related to arterial influence, the component after the dicrotic notch to the venous pressure. P2 ends on the dicrotic notch and P3 follows the dicrotic notch [30]. It is therefore likely that addition of the MAP signal to the ICP signal during expert labelling will reduce the inconsistency in ICP sub-peak labelling. Revised preprocessing might also improve the performance of the experts. The experts were a few times confronted with a non-physiological wave, causing extra confusion in the sub-peak assignment.

Spectral Regression

The spectral regression approach to detect ICP sub-peaks has been exploited in literature. The implementation of machine learning techniques improved the accuracy of peak detection compared to older techniques [18]. Scalzo et al obtained an 99% accuracy for peak designation [38, 39]. We implemented a version of the spectral regression method and obtained an accuracy of 78% for individual (leave one out) training and 91% for global (80-20) training.

The inconsistency in peak labelling by the experts is important to bear in mind when interpreting the results of the spectral regression algorithm. The algorithm was trained by ICP waves with sub-peaks labelled by experts. The results of the inter-and-intraclass correlation coefficient and the percentage of inconsistent labelling indicate a large variability in the training and test labels. The algorithm used the test and training set constructed by the experts as its gold standard. The inconsequent peak labels in the training set will learn the algorithm to label peak in the same inconsequent way. Due to the inconsistent labelling the peak assignment was erratic, and the

accuracy, sensitivity and specificity of the algorithm were low even though the performance was quite promising. The low accuracy, sensitivity and specificity were due to comparison with a 'gold standard' in which a large inconsistency in the labelled sub-peaks exists.

The same cautiousness is required for interpretation of the peak prediction error. The peak prediction error was quite large and varied between 2.3 samples in P1 to 6.2 samples for P3 in the leave one out cross-validation method. The algorithm performed quite well, but in some cases the algorithm assigned false peaks. The findings of the prediction error might be somewhat limited and should be interpreted with caution. The error was calculated by the sample difference between the sub-peaks labelled by the expert and predicted by the KSR. The error was unfairly large in some situation. In an example situation, the expert assigned three peaks (P1, P2, P3) and the algorithm assigned two peaks assigned as P1 and P2. Similar named ICP sub-peaks will be subtracted, it might occur that the KSR peaks named as P1 and P2 corresponds to the expert's peaks P2 and P3. This resulted occasionally in a large error, we did not correct for these large errors. In future research, it might be useful to ask the expert to assign a label (P1, P2, P3) to each of the labelled ICP-subpeaks. This will give more insight in the inconsistency in sub-peak labelling and will be helpful to calculate a more realistic prediction error.

The results of the prediction error, sensitivity, specificity and accuracy show the different performance of the 80-20 and individual training method. As expected, the 80-20 algorithm performed better than the individual training approach. Data from the same patients appeared both in the training and test set in the 80-20 approach. To have a viable algorithm it should perform well on previously unseen patients. According to the results of the individual training approach the performance of the algorithm is limited.

The performance of the spectral regression algorithm can be improved in future. Concluding from this discussion we have three important future recommendations. The dataset of 31 patients should be expanded. In this research a small amount of training and test data, only 730 labelled waves, were used. Elaboration of the data and the test and training set is an important issue for future research. A final important improvement is consistent peak labelling by experts. As the algorithm uses the expert labels as gold standard, it is important to have a reliable gold standard with consistent labels.

MOCAIP

A general important discussion point is the reliability of the MOCAIP metrics. The peaks predicted by the spectral regression method were used to calculate the metrics. As discussed before, the peak prediction is not perfect due to a limited dataset, inconsequent peak labelling and limitations

in the signal preprocessing. Interpretation of the results should be performed regarding this inaccuracy.

However, the results showed significance in 20 of the 27 metrics between the IH, pre-IH and control waves. This suggests the potential to classify based on these metrics. Many metrics were significant between IH and the pre-IH classes. The significance was limited between the pre-IH classes. This might indicate that the predictive power of morphological changes to predict IH is limited to a short period prior to IH. At least 5 metrics were significant between the control classes and the pre-IH waves, two of them were the mICP and diastolic ICP. The significant differences between waves leading to hypertension and control waves supports the hypothesis of development of morphological changes prior to hypertension. From our results, development of morphological changes in the period from 20 to 5 minutes prior to IH seemed limited, as there were no significant metrics between these periods, except for the mICP.

Classification by a classification tree - MOCAIP

However, the fact that a metric had no predictor importance does not signify the metric is not useful for classification. This can be illustrated by the significance of several metrics, which were not selected in the classification tree. This could be explained by reciprocal correlation with other metrics, as a result of the correlation it might not be selected for classification. Vice versa, metrics assessed as not significant by the Kruskal Wallis can be important for classification. The statistical analysis focused on the importance of one metrics, while the classification emphasizes the predictive power of a combination of metrics [10].

The experiment achieved a high specificity 96%, but a moderate sensitivity of 66% for classification of IH – waves and no – IH waves. Classification of the averaged waves in pre-IH and IH and control classes was quite promising.

We obtained promising results for IH prediction while classifying the IH, pre-IH and control waves. We could distinguish ICP waves leading to hypertension (within 20 minutes) from control waves with a sensitivity of 89%, specificity of 71%, accuracy of 86%. As the mICP and diasICP were excluded the sensitivity, specificity and accuracy were 74%, 66% and 73%. The predictive power of classification of pre-IH segments at a specific time prior to IH is still limited from this experiment. The clinical relevance of classification of waves leading to hypertension from control waves is high. This provides valuable information to clinician, as it allows to clinician to intervene before the pressure increases. From these results the clinician could be warned that a rise in ICP will occur within a period of 20 minutes.

It is important to realize that control segments were selected from episodes with no intracranial hypertension for a minimum of 90 minutes over the full length of the ICP signal. Control segments could be selected before and after the IH period. Morphological alterations might have occurred before the control segment in previous IH periods. In this cases discrimination between IH, pre-IH and control might be impossible as the morphological changes have yet occurred in all segments. It is recommended to select control segments prior to the period of increasing hypertension and verify if it is possible to classify these control waves from the waves leading to hypertension.

Criteria for segment selection were based on the ICP guidelines and previous literature [9, 10]. An IH segment was selected as the mean pressure was above 22 mmHg for a minimum of 5 consecutive minutes, detected by a sliding window averaging over 1 minute. Further studies require evaluation of this selection protocol. Segments might be missed due to the strong selection criteria.

Classification by a Neural Network

The relevance of ICP wave classification by a neural network is limited from this research. We obtained better classification results based on the MOCAIP metrics. Beside the advantage of better classification results, the classification tree based on MOCAIP metrics has the benefit to be very insightful. The classification tree is similar to a clinical protocol. The neural network is a black box, which complicates acceptance by the clinicians. The neural network performances might benefit improved preprocessing. In future, it might be interesting to explore classification based on the MOCAIP metrics by a neural network. As Matlab could not deal NaN values during classification by a neural network we could not elaborate on this performance. Another interesting focus is the use of LSTM Neural Networks to forecast the ICP values of future timesteps. As the values of multiple time step could be forecasted, it would be possible to predict IH.

Clinical application

Although our results need further study, a future application of the algorithm is to process continuous ICP signals and extract MOCAIP metrics. Bedside analysis of the MOCAIP metrics would provide more information than the mean ICP to the clinician. The algorithm should be adapted to allow real-time analysis of the signal and monitoring of the most important MOCAIP metrics. This would be valuable to provide predictive information that warns for increasing intracranial pressure and progression of a secondary insult. It is important to be aware of the risk of false positive predictions or failure to predict intracranial hypertension.

Conclusion

Our aim was to predict intracranial hypertension. We presented two machine learning methods to classify IH waves from no-IH waves. The MOCAIP method based on a spectral regression algorithm obtained promising results for classification between IH and no-IH waves. The classification results of our second approach based on neural networks, using the full wave as its input, were limited. Both methods might benefit improved pre-processing and expansion of the training, test and dataset. The MOCAIP method was elaborated to predict intracranial hypertension. It was feasible to classify between ICP waves leading to hypertension and control waves based on the MOCAIP metrics. These results are promising for future clinical practice as it could provide valuable information to the clinician, allowing to intervene within a period of 20 minutes before intracranial hypertension.

References

- [1] Z. Idris, M. Mustapha, and J. M. Abdullah, *Neurointensive Care Monitoring for Severe Traumatic Brain Injury*. 2012.
- [2] M. Aboy, J. McNames, and B. Goldstein, "Automatic detection algorithm of intracranial pressure waveform components," *Conf. Proc. 23rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 3, pp. 2231–2234, 2001.
- [3] B. Mokri, "The Monro–Kellie hypothesis Applications in CSF volume depletion," *Neurology*, vol. 56, pp. 1746–1748, 2001.
- [4] W. Peeters, R. van den Brande, S. Polinder, and A. Brazinova, "Epidemiology of traumatic brain injury in Europe," *Acta Neurochir. (Wien)*, vol. 157, no. 10, pp. 1683–1696, Oct. 2015.
- [5] P. Le Roux, D. Menon, G. Citerio, and P. Vespa, "Consensus Summary Statement of the International Multidisciplinary Consensus Conference on Multimodality Monitoring in Neurocritical Care," *Neurocrit. Care*, vol. 21, no. S2, pp. 1–26, Dec. 2014.
- [6] K. Wartenberg, J. Schmidt, and S. Mayer, "Multimodality Monitoring in Neurocritical Care," *Crit. Care Clin.*, vol. 23, no. 3, pp. 507–538, 2007.
- [7] L. Steiner and P. Andrews, "Monitoring the injured brain: ICP and CBF," *Br. J. Anaesth.*, vol. 97, no. 1, pp. 26–38, Jul. 2006.
- [8] P. Andrews and G. Citerio, "Intracranial pressure Part one: Historical overview and basic concepts," *Intensive Care Med*, vol. 30, pp. 1730–1733, 2004.
- [9] Brain Trauma Foundation, "Guidelines for the Management of Severe Traumatic Brain Injury 4th Edition," 2016.
- [10] X. Hu, P. Xu, S. Asgari, P. Vespa, and M. Bergsneider, "Forecasting ICP Elevation Based on Prescient Changes of Intracranial Pressure Waveform Morphology," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 5, pp. 1070–1078, May 2010.
- [11] P. Le Roux, *Intracranial Pressure Monitoring and Management*. CRC Press/Taylor and Francis Group, 2016.
- [12] P. Talving, E. Karamanos, P. Teixeira, G. Skiada *et al.*, "Intracranial pressure monitoring in severe head injury: compliance with Brain Trauma Foundation guidelines and effect on outcomes: a prospective study," *J. Neurosurg.*, vol. 119, no. 5, pp. 1248–1254, Nov. 2013.
- [13] A. Farahvar, L. Gerber, Y. Chiu, N. Carney, R. Härtl, and J. Ghajar, "Increased mortality in patients with severe traumatic brain injury treated without intracranial pressure monitoring," *J. Neurosurg.*, vol. 117, no. 4, pp. 729–734, Oct. 2012.
- [14] A. Alali, R. Fowler, T. Mainprize, and D. Scales, "Intracranial Pressure Monitoring in Severe Traumatic Brain Injury: Results from the American College of Surgeons Trauma Quality

- Improvement Program," *J. Neurotrauma*, vol. 30, no. 20, pp. 1737–1746, Oct. 2013.
- [15] R. Hamilton, X. Peng, S. Asgari, M. Kasprowicz *et al.*, "Forecasting intracranial pressure elevation using pulse waveform morphology," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009, pp. 4331–4334.
- [16] A. Calisto, A. Bramanti, M. Galeano, and F. Angileri, "A preliminary study for investigating idiopathic normal pressure hydrocephalus by means of statistical parameters classification of intracranial pressure recordings," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009, pp. 2629–2632.
- [17] E. Carrera, D. Kim, G. Castellani, C. Zweifel *et al.*, "What Shapes Pulse Amplitude of Intracranial Pressure?," *J. Neurotrauma*, vol. 27, no. 2, pp. 317–324, Feb. 2010.
- [18] X. Hu, P. Xu, F. Scalzo, P. Vespa, and M. Bergsneider, "Morphological Clustering and Analysis of Continuous Intracranial Pressure," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 3, pp. 696–705, 2009.
- [19] C. Nucci, P. De Bonis, A. Mangiola, P. Santini *et al.*, "Intracranial pressure wave morphological classification: automated analysis and clinical validation," *Acta Neurochir.*, vol. 158, no. 3, pp. 581–588, Mar. 2016.
- [20] R. Myers, C. Lazaridis, C. Jermaine, C. Robertson, and C. Rusin, "Predicting Intracranial Pressure and Brain Tissue Oxygen Crises in Patients With Severe Traumatic Brain Injury," *Crit. Care Med.*, vol. 44, no. 9, pp. 1754–1761, Sep. 2016.
- [21] F. Güiza, B. Depreitere, I. Piper, G. Van den Berghe, and G. Meyfroidt, "Novel Methods to Predict Increased Intracranial Pressure During Intensive Care and Long-Term Neurologic Outcome After Traumatic Brain Injury," *Crit. Care Med.*, vol. 41, no. 2, pp. 554–564, Feb. 2013.
- [22] F. Scalzo and X. Hu, "Semi-supervised detection of intracranial pressure alarms using waveform dynamics," *Physiol. Meas.*, vol. 34, no. 4, pp. 465–478, Apr. 2013.
- [23] F. Scalzo, D. Liebeskind, and X. Hu, "Reducing False Intracranial Pressure Alarms Using Morphological Waveform Features," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 1, pp. 235–239, Jan. 2013.
- [24] F. Scalzo, S. Asgari, S. Kim, M. Bergsneider, and X. Hu, "Bayesian tracking of intracranial pressure signal morphology," *Artif. Intell. Med.*, vol. 54, no. 2, pp. 115–123, Feb. 2012.
- [25] B. Quachtran, R. Hamilton, and F. Scalzo, "Detection of Intracranial Hypertension using Deep Learning," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, 2016, pp. 2491–2496.
- [26] R. Ryan Morton, "Intracranial Hypertension | Clinical Gate," *13-03-2015*, 2015..

- [27] P. Eide, "The correlation between pulsatile intracranial pressure and indices of intracranial pressure-volume reserve capacity: results from ventricular infusion testing.," *J. Neurosurg.*, vol. 125, pp. 1493–1503, 2016.
- [28] F. Scalzo, M. Bergsneider, P. Vespa, N. Martin, and X. Hu, "Intracranial Pressure Signal Morphology: Real-Time Tracking," *IEEE Pulse*, vol. 3, no. 2, pp. 49–52, Mar. 2012.
- [29] K. Evensen, M. O'Rourke, F. Prieur, S. Holm, and P. Eide, "Non-invasive Estimation of the Intracranial Pressure Waveform from the Central Arterial Blood Pressure Waveform in Idiopathic Normal Pressure Hydrocephalus Patients," *Sci. Rep.*, vol. 8, no. 1, p. 4714, Dec. 2018.
- [30] C. Kirkness, P. Mitchell, R. Burr, and K. March, "Intracranial pressure waveform analysis: Clinical and research implications," *J. Neurosci. Nurs.*, vol. 32, no. 5, p. 271, 1986.
- [31] W. Armstead, "Cerebral Blood Flow Autoregulation and Dysautoregulation.," *Anesthesiol. Clin.*, vol. 34, no. 3, pp. 465–77, Sep. 2016.
- [32] M. Czosnyka and J. Pickard, "Monitoring and interpretation of intracranial pressure.," *J. Neurol. Neurosurg. Psychiatry*, vol. 75, no. 6, pp. 813–21, Jun. 2004.
- [33] R. Bray, A. Sherwood, J. Halter, C. Robertson, and R. Grossman, "Development of a Clinical Monitoring System by Means of ICP Waveform Analysis," in *Intracranial Pressure VI*, Berlin, Heidelberg: Springer Berlin Heidelberg, 1986, pp. 260–264.
- [34] F. Scalzo, R. Hamilton, and X. Hu, "Real-Time Analysis of Intracranial Pressure Waveform Morphology," in *Advanced Topics in Neurological Disorders*, K.-S. Chen, Ed. Rijeka: IntechOpen, 2012.
- [35] F. Scalzo, R. Hamilton, S. Asgari, S. Kim, and X. Hu, "Intracranial hypertension prediction using extremely randomized decision trees.," *Med. Eng. Phys.*, vol. 34, no. 8, pp. 1058–65, Oct. 2012.
- [36] P. Shrout and J. Fleiss, "Intraclass correlations: uses in assessing rater reliability.," *Psychol. Bull.*, vol. 86, no. 2, pp. 420–8, Mar. 1979.
- [37] T. Koo and M. Li, "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research.," *J. Chiropr. Med.*, vol. 15, no. 2, pp. 155–63, Jun. 2016.
- [38] F. Scalzo, P. Xu, M. Bergsneider, and X. Hu, "Nonlinear regression for sub-peak detection of intracranial pressure signals," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008, pp. 5411–5414.
- [39] F. Scalzo, P. Xu, S. Asgari, M. Bergsneider, and X. Hu, "Regression analysis for peak designation in pulsatile pressure signals," *Med. Biol. Eng. Comput.*, vol. 47, no. 9, pp. 967–977, Sep. 2009.
- [40] D. Cai, X. He, and J. Han, "Spectral regression for dimensionality reduction," 2007.
- [41] D. Cai, X. He, and J. Han, "Speed up kernel discriminant analysis," *VLDBJ.*, vol. 20, no. 1, pp.

21–33, Feb. 2011.

- [42] D. Cai, X. He, W. V. Zhang, and J. Han, “Regularized locality preserving indexing via spectral regression,” in *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management - CIKM '07*, 2007, p. 741.
- [43] R. Storn, *On the usage of differential evolution for function optimization*. IEEE, 1996.
- [44] E. Cardoso, J. Rowan, and S. Galbraith, “Analysis of the cerebrospinal fluid pulse wave in intracranial pressure,” *J. Neurosurg.*, vol. 59, no. 5, pp. 817–821, Nov. 1983.
- [45] H. Lee, E. Jeong, H. Kim, and M. Cazosnyke, “Morphological Feature Extraction From a Continuous Intracranial Pressure Pulse via a Peak Clustering Algorithm,” *IEEE Trans. Biomed. Eng.*, vol. 63, no. 10, pp. 2169–2176, Oct. 2016.

Appendices

Appendix 1 – Spectral Regression Algorithm

We have a matrix X of input waves $[x_1 \dots x_n]$. Each vertex represents a wave.

Let W be a symmetric $m * m$ matrix with W_{ij} having the weight of the edge joining vertices. The purpose of graph embedding is to represent each vertex of the matrix as a low dimensional vector preserving the similarities between the vertex pairs.

Let $y = [y_1, y_2, \dots, y_m]^T$ be the map from the matrix to the real line. The optimal y can be found by minimizing

$$\sum_{i,j} (y_i - y_j)^2 W_{ij}, \quad (\text{A1})$$

$$\sum_{i,j} (y_i - y_j)^2 W_{ij} = 2y^T L y. \quad (\text{A2})$$

Where W is the affinity matrix, $L = D - W$ is the graph Laplacian and D is the diagonal matrix of W whose entries are the column sums of W .

The optimal y 's can be obtained by solving the maximum eigen value problem:

$$W y = L D y. \quad (\text{A3})$$

This graph embedding approach describes the mapping of the training set. As we would like to classify the ICP sub-peak in previously unseen ICP waves. We can choose a linear function

$$y_i = f(x_i) = a^T x_i \text{ we have } y = X^T a. \quad (\text{A4})$$

The optimal a 's are the eigenvectors corresponding to the eigen value problem:

$$X W X^T a = \lambda X D X^T a. \quad (\text{A5})$$

Find a which satisfies $y = X^T a$.

A possible way to find a is via a regularized least squares problem,

$$a = \operatorname{argmin}_\alpha \sum_{i=1}^n (\alpha^T x_i - y_i) + \delta \|\alpha\|^2. \quad (\text{A6})$$

The regularization problem can be rewritten as:

$$\alpha = (X X^T + \delta I)^{-1} X^T y. \quad (\text{A7})$$

Where X is a matrix formed by the input waves $[x_1 \dots x_n]$, α is the eigenvector, I is the identity matrix and δ is the regularization parameters, we used $\delta = 0.001$. This formula can be solved using Cholesky decomposition [21].

We used Kernel spectral regression to extend this method to nonlinear problems. This allows to use a linear regression analysis to solve a nonlinear problem. The kernel maps the input waves x a higher dimensional space. We used a Gaussian kernel (K):

$$K_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right). \quad (\text{A8})$$

Where sigma is the standard deviation of the kernel. Similar to the spectral regression approach we use Cholesky decomposition to obtain the vector α ,

$$r = \text{chol}(K + \delta I) \quad (\text{A9})$$

$$\alpha = r \backslash (r^T \backslash y). \quad (\text{A10})$$

Now we can find the embedding function:

$$\text{Embedding function} = K * \alpha \quad (\text{A11})$$

where α is the eigen vector that minimized the residual sum of square error in equation A6.

We calculated the class centers (mean of the embedding functions) of the wave with the same label in the training set [42].

The Kernel of the test set was constructed, in a similar way as the Kernel from the training set, equation A8. The embedding function of the test set was calculated by equation A11. The corresponding labels were assigned by minimization of the Euclidian distance between the embedding functions of the test set and the class centers of the embedding functions of the training set. The labels of the class center that minimized the Euclidian distance were assigned to the waves in the test set. The KSR algorithm assigned 1, 2, or 3 ICP-subpeaks to the averaged ICP waves.

Appendix 2 – MOCAIP metrics

- Amplitude: The amplitudes dP1, dP2 and dP3 are the height of the peaks from the diastolic ICP height to the top of the peaks. The diastolic ICP was determined as the minimal value in the averaged ICP-pulse. (dP2/dP1), (dP3/dP1) and (dP3/dP2) were the ratios of the peak amplitudes.
- Latency: The peak distances in horizontal direction were (L1, L2 and L3). It was the x-coordinate of the diastole. The difference in latencies were calculated as L2-L1, L3-L1 and L3-L2.
- Curvature: The curvatures (curv1, curv2, curv3) of the peaks were calculated by the following formula A12:

$$Curv(n) = \frac{|x''(n)|}{(1+x'(n)^2)^{3/2}}. \quad (A12)$$

Where x is the averaged wave and n corresponds to the position (1,2, or 3) of the peak. Curm is the mean of the curvatures.

- Lx: is the position where the signal returned to 37% of the height of P3.
- Slope: The slope was calculated by the following formula:

$$Slope = \frac{P1-diasICP}{Lp1}. \quad (A13)$$

Two extra metrics were added compared to the original MOCAIP algorithm. These metrics were added as they did not include the position of P1, P2 or P3. They are robust for limitations of the spectral regression algorithm.

- Amplitude: The amplitude was the difference between the diastolic ICP and the max height of the wave. This metric represents the hypothesis that a raised ICP amplitude could be indicative of impaired intracranial compliance and pressure-buffering capacity of the brain [27].
- Max Slope: The max slope was the max amplitude of the wave divided by the latency of the diastolic ICP to the point where the ICP was maximal.

Which metrics were calculated depended on the peaks assigned by the regression model. Metrics were assigned empty as they were based on a non-existent subpeak.

Appendix 3 – Interrater correlation coefficient and percentage of inconsistent labelling

Table A1: The intra-rater reliability for all seven experts who labelled peaks in single ICP waves. The intraclass correlation coefficient was calculated for P1, P2, P3. Inconsistent labelling P1, P2 or P3 was the percentage of waves with inconsistency between and the first and second time the expert assigned a label to the wave.

Rater	Interclass correlation, P1	Interclass correlation, P2	Interclass correlation, P3	Inconsistent labelling P1 (%)	Inconsistent labelling P2 (%)	Inconsistent labelling P3 (%)
1	0.999 (0.993 - 1)	0.998 (0.993 - 1)	1 (0.999 - 1)	10%	0%	10%
2	1 (0.997 - 1)	0.997 (0.990- 0.999)	1 (0.999 - 1)	10%	0%	10%
3	0.999 (0.997 - 1)	1 (0.999 - 1)	1 (0.999 - 1)	10%	0%	0%
4	0.998 (0.991 - 1)	0.999 (0.998 - 1)	1 (0.999 - 1)	10%	0%	20%
5	1 (0.999 - 1)	0.998 (0.994 - 1)	1 (0.999 - 1)	10%	0%	0%
6	1 (0.998 - 1)	0.999 (0.996 - 1)	0.998 (0.983 - 1)	0%	0%	0%
7	0.999 (0.989 - 1)	0.999 (0.997 - 1)	1 (1 - 1)	0%	0%	10%

Appendix 4 – Performance of spectral regression algorithm

The kernel spectral regression algorithm was trained by a 31 leave on out training approach. We explored the effect of a different training approach by a random selection of waves in the test and training set. 80% was used for training and 20% to test the performance. This improved the performance compared to the leave one out training approach. This improvement was expected as the training and test included waves from the same patients.

The prediction error for P1 was 0.93 samples, P2 was 2.5 samples and the error for P3 was 3.2 samples. The average prediction error was 2.2 samples.

The prediction error was the error calculated from a test and training set with 80% of the labelled waves in the training set and 20% in the test set. The sensitivity, specificity and accuracy for P1, P2 and P3 can be observed in table 4.

Table A2: Sensitivity, specificity and accuracy for assignment of the subpeaks P1, P2 and P3 predicted by the 80-20 training method. 20% of the data was used in the test set and 80% in the training set.

Peak	Sensitivity	Specificity	Accuracy
P1	87%	77%	82%
P2	100%	-	100%
P3	98%	60%	92%
P123	97%	73%	91%

Appendix 5 – MOCAIP statistics

Table A3: Summery statistics of the Kruskal Wallis test. The green color indicated statistical significance between the two groups for the corresponding metric. Yellow indicates not statistical significance between the two groups. The first column is the P-value, a P-value <0.05 states a significant change over time for the corresponding metric in the period from control to intracranial hypertension.

	P-value	IH vs 5	IH vs 10	IH vs 15	IH vs 20	IH vs Control	5 vs 10	5 vs 15	5 vs 20	5 vs Control	10 vs 15	10 vs 20	10 vs Control	15 vs 20	15 vs Control	20 vs Control
mICP	<0.0001															
diasICP	<0.0001															
dP1	<0.0001															
dP2	<0.0001															
dP3	<0.0001															
Ratiod12	0.0028															
Ratiod13	< 0,0001															
Ratiod23	< 0,0001															
Amplitude	< 0,0001															
Lt	0.0007															
L1	0.9282															
L2	0.8573															
L3	0.0005															
RatioL12	0.6063															
RatioL13	0.0014															
RatioL23	0.0007															
Curv1	0.9169															
Curv2	< 0,0001															
Curv3	< 0,0001															
Curvm	< 0,0001															
Curv12	0.1956															
Curv31	0.3033															
Curv32	0.0004															
Slope	< 0,0001															
Lx	0.1362															
dPx	< 0,0001															
Slopemax	< 0,0001															

