Forecasting intracranial hypertension using multi-scale waveform metrics

Matthias Hüser, Adrian Kündig, Walter Karlen, Valeria De Luca, and Martin Jaggi

***Abstract*—Objective: Acute intracranial hypertension is an important risk factor of secondary brain damage after traumatic brain injury. Hypertensive episodes are often diagnosed reactively and time is lost before counteractive measures are taken. A proactive approach that predicts critical events several hours ahead of time could be beneficial for the patient. Methods: We developed a prediction framework that forecasts onsets of acute intracranial hypertension in the next 8 hours. It jointly uses cerebral autoregulation indices, spectral energies and morphological pulse metrics to describe the neurological state of the patient. One-minute base windows were compressed by computing signal metrics, and then stored in a multi-scale history, from which physiological features were derived. Results: Our model predicted events up to 8 hours in advance with alarm recall rates of 90% at a precision of 36% in the MIMIC-II waveform database, improving upon two baselines from the literature. We found that features derived from high-frequency waveforms substantially improved the prediction performance over simple statistical summaries of low-frequency time series, and each of the three feature classes contributed to the performance gain. The inclusion of long-term history up to 8 hours was especially important. Conclusion: Our approach showed promising performance and enabled us to gain insights about the critical components of the prediction model. Significance: Our results highlight the importance of information contained in high-frequency waveforms in the neurological intensive care unit. They could motivate future studies on pre-hypertensive patterns and the design of new alarm algorithms for critical events in the injured brain.**

***Index Terms*—Cerebral auto-regulation indices, Intracranial hypertension, Intracranial pressure, Machine learning, ICP pulse morphology**

I. Introduction

With at least 10 million cases annually leading to hospitalization worldwide, traumatic brain injury (TBI), often causing intracranial hemorrhage, is a major public health issue [1]. After initial admission to the intensive care unit (ICU) and assessment of the primary brain injury, further neurological damage often occurs. This phenomenon is referred to as *secondary brain injury*, and often leads to long-term brain damage through e.g. cerebral ischemia (decrease of blood flow to the brain) [2], cerebral hypoxia (decrease of substrate/oxygen flow to the brain) [3] and brain herniation (swelling leading to compression of brain structures [4]).

Management of TBI patients in the neurological ICU focuses on mitigating and possibly reversing secondary injuries [5]. A key variable in the management of secondary brain injury is *intracranial pressure* (ICP) [6, 7]. Cerebral compliance maintains blood- and energy substrate flow by holding pressure constant against slight volume changes of the cranial components [8]. The ICP value of a healthy adult is maintained by this mechanism in the range 7-15 mmHg [9]. However, if compliance is reduced, rapid non-linear ICP elevations can occur [10]. A sustained ICP elevation over 20 mmHg is defined as *acute intracranial hypertension* (ICH) [11]. An illustrative example of an ICH event is shown in Fig. 1.

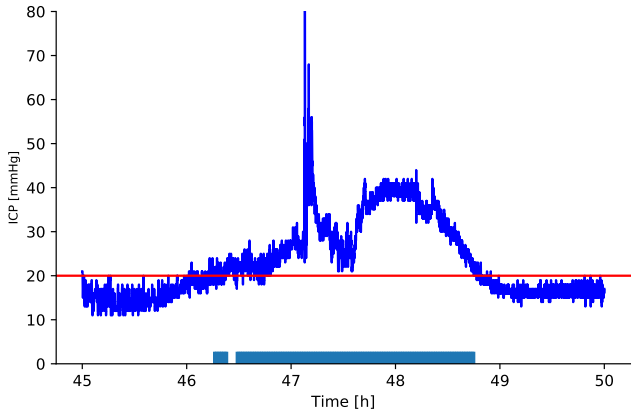

Fig. 1. ICP recording illustrating an acute intracranial hypertension event. The patient exhibits normal ICP values, before deteriorating into intracranial hypertension (ICH, boxes on rug plot). At the end of the trace the patient returns to a normal ICP state. The horizontal red line denotes the threshold defining ICH.

A direct association of time spent in the ICH state with clinical outcome has been empirically shown: The area under the ICP curve in the first 48 hours of ICU treatment is an independent predictor of in-hospital mortality [12]. Various other studies have established an association of ICH and poor neurological outcome [13]–[15]. Accordingly, it is a common treatment goal in neuro-critical care to avoid acute intracranial hypertension [7]. Invasive, intra-parenchymal ICP monitoring combined with interventions such as external ventricular drainage or surgery is the gold standard to control and maintain

This work was supported by the Gebert-Rüf Stiftung, Switzerland, under grant agreement GRS-025/14 and the Swiss National Science Foundation, under grant agreement 150640. MH was partially funded by the Grant No. 205321_176005 "Novel Machine Learning Approaches for Data from the Intensive Care Unit" of the Swiss National Science Foundation (to Gunnar Rätsch).

M. Hüser is with the Biomedical Informatics Group, Department of Computer Science, ETH Zürich, Universitätstrasse 6, 8092 Zürich, Switzerland (correspondence e-mail: matthias.hueser@inf.ethz.ch). A. Kündig was with the Department of Computer Science, ETH Zürich, 8092 Zürich, Switzerland. W. Karlen is with the Mobile Health Systems Lab, Institute of Robotics and Intelligent Systems, Department of Health Sciences and Technology, ETH Zürich, 8008 Zürich, Switzerland. V. De Luca was with the Department of Information Technology and Electrical Engineering, ETH Zürich, 8092 Zürich, Switzerland. She is now with the Novartis Institutes for Biomedical Research, 4056 Basel, Switzerland. M. Jaggi is with the Machine Learning & Optimization Lab, EPFL, 1015 Lausanne, Switzerland.

ICP in the physiological range of 7-15 mmHg and ensure adequate cerebral compliance [9]. Advances in monitoring and signal processing technology have allowed to record high-frequency ICP traces and analyze them in real-time [16]. Yet there are several caveats that hinder the interpretation of the ICP signal and its use as a decision-support tool: (a) raw data and time-varying trends are presented to the clinician, and no risk estimates for ICH are available. This can lead to information overload and over-consumption of human attention for the ICU personell. For example, a study has found that clinicians are often not confident that effort spent on inspection of ICP traces is redeemed by improving outcome after TBI [17]; (b) threshold-based track-and-trigger systems usually have too high false alarm rates, which can desensitize staff to dangerous hypertension events [18]; (c) alarms are only triggered after onset of acute intracranial hypertension, when long-term effects might be harder to prevent.

To address these problems, robust *forecasting* of ICH onsets could augment the current treatment protocol which is overly manual, reactive, and prone to errors. Previous works have shown that complex precursor patterns occur in auto-regulation indices and ICP/ABP waveform morphology prior to hypertensive events [19, 20]. Recently, simple prediction models explicitly targeting ICH forecasting 30 minutes up to 6 hours before the event were proposed, yielding promising results [21]–[23]. However, it is not well understood or investigated which of these approaches is necessary or sufficient to achieve high prediction performance in the context of an early warning system for ICH, and whether additional benefits could be derived from their combination.

In this extensive empirical study of ICH prediction we make the following contributions

- An online ICH prediction framework which describes the neurological state using multi-scale metrics of the last 8 hours of recordings, comprising classical statistical features, cerebral auto-regulation indices, frequency band energies and ICP/ABP pulse morphology computed on high-frequency waveforms. The resulting model is shown to outperform two baseline models from the literature in a controlled comparison.
- By including a wide range of relevant channels and physiological feature types, we conduct the first systematic study of ICH prediction across signal channels and feature types, and thereby benchmark various pre-hypertensive patterns exploited or hypothesized in previous works.
- We demonstrate clear performance benefits when including morphological and spectral energy features derived from high-frequency waveforms compared to focusing on only statistical metrics on low-frequency time series which have been often used in major recent works.
- Using the state-of-the-art feature attribution technique SHAP (SHapley Additive exPlanation) we study the importance and generate rankings of different features that explain positive intracranial hypertension alarms, representing the first application of this technique to an extensive set of pre-hypertensive patterns.

Preliminary and partial versions of this work have been reported in clinical abstracts [24, 25].

II. Related work

The association of information contained in high-frequency physiological waveforms/time series and elevated ICP has been studied in various works. For example, Hornero et al. [26] have found that decreased ICP signal complexity and irregularity is associated with intracranial hypertension. Fan et al. [27] identified an association between ICP variability and decreased pressure auto-regulation. Recently, it was established that characteristic patterns in various physiological channels are correlated with ICP and could thus be used to predict ICH [28]. Several auto-regulation indices defined on physiological channels were reported, such as by Zeiler et al. [29], which studied the moving correlation coefficient between ICP, ABP and CPP channels, and others [30, 31]. The relationship between auto-regulation indices and successive ICH events has been studied by Kim et al. [32]. In general, it has long been suspected that the information contained in the pulsatile ICP signal is very rich beyond simple statistical summaries [33, 34].

Besides auto-regulation indices, previous works have attempted to use morphological descriptors of the intracranial pressure pulse to predict ICH onset up to 20 minutes in advance [20, 35, 36]. More generally, morphological analysis of ICP pulses [37] has emerged as a successful approach and was used to e.g. reduce false alarm rates of ICP alarms [38] and track pulse metrics in real-time [39]. Hu et al. [40] applied cluster analysis to individual ICP pulses. Other types of features that have been proposed to summarize physiological time series include bag-of-words of physiological motifs applied to ECG/EEG time series [41] and entropy measures [42, 43]. The recently proposed ICP trajectories framework [44] uses longitudinal ICP time series to discover clinical phenotypes. Different approaches have also been proposed, based on assessing risk only from static clinical data [45] or biomarkers [46]–[48], instead of using historical time series.

To obviate the need for explicit feature engineering on historical time series, deep learning architectures have been proposed, which detect intracranial hypertension from the raw pulse waveform [49]. Simpler dimensionality reduction approaches, such as principal component analysis, have also been used to find non-correlated features [50] that describe ICH.

Major recent works explicitly addressing the ICH forecasting problem include the approach proposed by Güiza et al. [22], which obtained an AUROC of 0.87 for prediction of ICH in the next 30 minutes. Their analysis showed that the most predictive channel is ICP and that the most recent measurements are the most relevant features. Subsequently, their model was externally validated, resulting in similar performance [23]. Myers et al. [21] proposed a model that is able to predict ICH up to 6 hours in advance, a prediction horizon comparable to our method. It uses simple features such as the last measured ICP value or the time to the last ICH crisis. Besides tackling the classification task directly, other

models have been suggested that predict the future ICP mean value, for example by using nearest-neighbor regression [51], neural networks [52, 53] or ARIMA models [54].

III. Methods

A. Physiological database

In all experiments, we have used the multi-parameter intelligent monitoring in intensive care II waveform database (MIMIC-II WFDB) [55], Version 3.2. The entire dataset consists of 17,468 adult patient stays at the Beth Israel Deaconess Medical Center, Boston, MA, United States. The MIMIC-II WFDB was chosen for this study because it contains simultaneous measurements of high-frequency waveforms (125 Hz) and derived time series (1 Hz) for a range of physiological channels that are relevant to the prediction problem. Among all available channels, we selected `ICP` (mean intracranial pressure) , `CPP` (cerebral perfusion pressure) and `ABPm/d/s` (mean/diastolic/systolic arterial blood pressures) time series, and `wICP` (intracranial pressure), `wABP` (invasive arterial blood pressure), `wPLETH` (raw output of fingertip plethysmograph), `wRESP` (respiration waveform) and `ECG` waveforms (Fig. 3a). This broad range allows us to compare the relevance of different channels for predictive modeling, while ensuring that we can extract a cohort of at least 50 ICU stays with regular sampling. Waveforms were acquired using the bedside IntelliVue Patient Monitoring system, Philips Healthcare, The Netherlands.

B. Cohort selection

Only a small fraction of available records in the MIMIC-II WFDB contain ICP data. In a first step, we discarded all segments that have no available `ICP` time series, which left 346 relevant segments. We further require a minimum recording length of 24 hours, and a missing value ratio of at most 25% for each considered waveform or time series channel. We applied these criteria to ensure that the relevance of different channels as features could be meaningfully compared, and individual channels would not be negatively affected by long stretches of missing data. After applying these criteria, 54 segments remained in the cohort. This set of recording segments was used in all reported experiments. A diagram summarizing patient exclusions and cohort definition is shown in Fig. 2. Matching to MIMIC-II clinical database records was only possible for 32 of the 54 segments, contributed by 22 unique patients, for which clinical context is provided in Table I. In terms of admission diagnosis, the cohort is homogeneous, with most patients exhibiting intracranial hemorrhage. In this work, we did not include clinical covariates into the processing pipeline to avoid unequal treatment of segments or a significant reduction of available segments. Overall, our data-set contains 4382 hours of data (Fig. 3a). Each segment has a mean recording length of 81.15 hours (std: 46.46 hours). The mean ICP value in the cohort is 9.79 mmHg (std: 7.5 mmHg).

C. Acute intracranial hypertension alarms

We define an ICH event as 5 successive 1-minute blocks with median ICP greater than 20 mmHg, which helps to avoid spurious labelings, following the recent work by Ziai et al. [56]. According to this definition, patients were in an acute ICH event state for 3.9% of their cumulative segment lengths. When analyzed by segment, the ICH state was active for a mean of 3.0% (std: 10.1%) of each segment's duration. A time point on the 1-minute grid was labeled as positive if there is *any acute intracranial hypertension* event in the next 8 hours and the patient is not already hypertensive (Fig. 3b). This strategy implements an early warning system deployed in phases where the patient has normal ICP values. Our design choice was to train one overall model for predicting events in the next 8 hours, without targeting any specific prediction horizon. Positive labels correspond to time points at which an alarm should be produced by our prediction model (Fig. 3b). Recall is defined as the fraction of those points, at which an alarm is indeed produced. Precision denotes the fraction of produced alarms which are in the 8 hours prior to some ICH event (Fig. 3g). Both metrics are maximized if continuous sequences of alarms are produced *exactly* in the 8 hours before events, one for each grid point. However, in clinical implementation this strict condition could be relaxed by applying post-processing such as moving window functions over the sequence of thresholded prediction scores. We consider such processing to be out-of-scope here, but we suspect it can improve practical alarm system performance significantly, both in terms of recall and false alarm rate. There were a total of 225,879 labeled 1-minute time points, of which 52,685 were

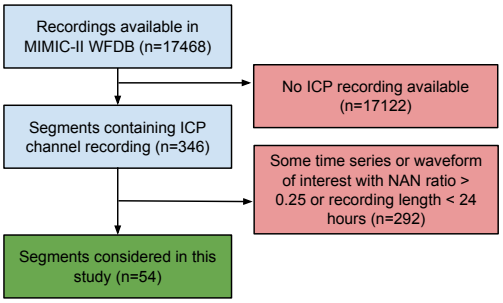

Fig. 2. Overview of selection criteria that were used to define the recordings of interest for this study

TABLE I
PATIENT DEMOGRAPHIC INFORMATION FOR ICU STAYS (n_s=22) MATCHED WITH $n = 32$ RECORDING SEGMENTS, $n = 22$ RECORDING SEGMENTS COULD NOT BE MATCHED IN THE MIMIC-II CDB

Age median [IQR]	64.5 [60-76]
Sex (% male)	50
Hospital mortality rate (%)	27.3
ICU LOS days median [IQR]	12.5 [6.4-20.7]
Admission type	Emergency ($n_s = 21$)
	Elective ($n_s = 1$)
ICU care service	Neurological surgical ($n_s = 19$)
	Neurological medical ($n_s = 2$)
	Cardiac medical ($n_s = 1$)
Diagnosis	Intracranial hemorrhage ($n_s = 10$)
	Subarachnoid hemorrhage ($n_s = 7$)
	Interparenchymal hemorrhage ($n_s = 2$)
	Bleed/head bleed ($n_s = 2$)
	Brain tumor ($n_s = 1$)
GCS median [IQR]	3 [3-4]

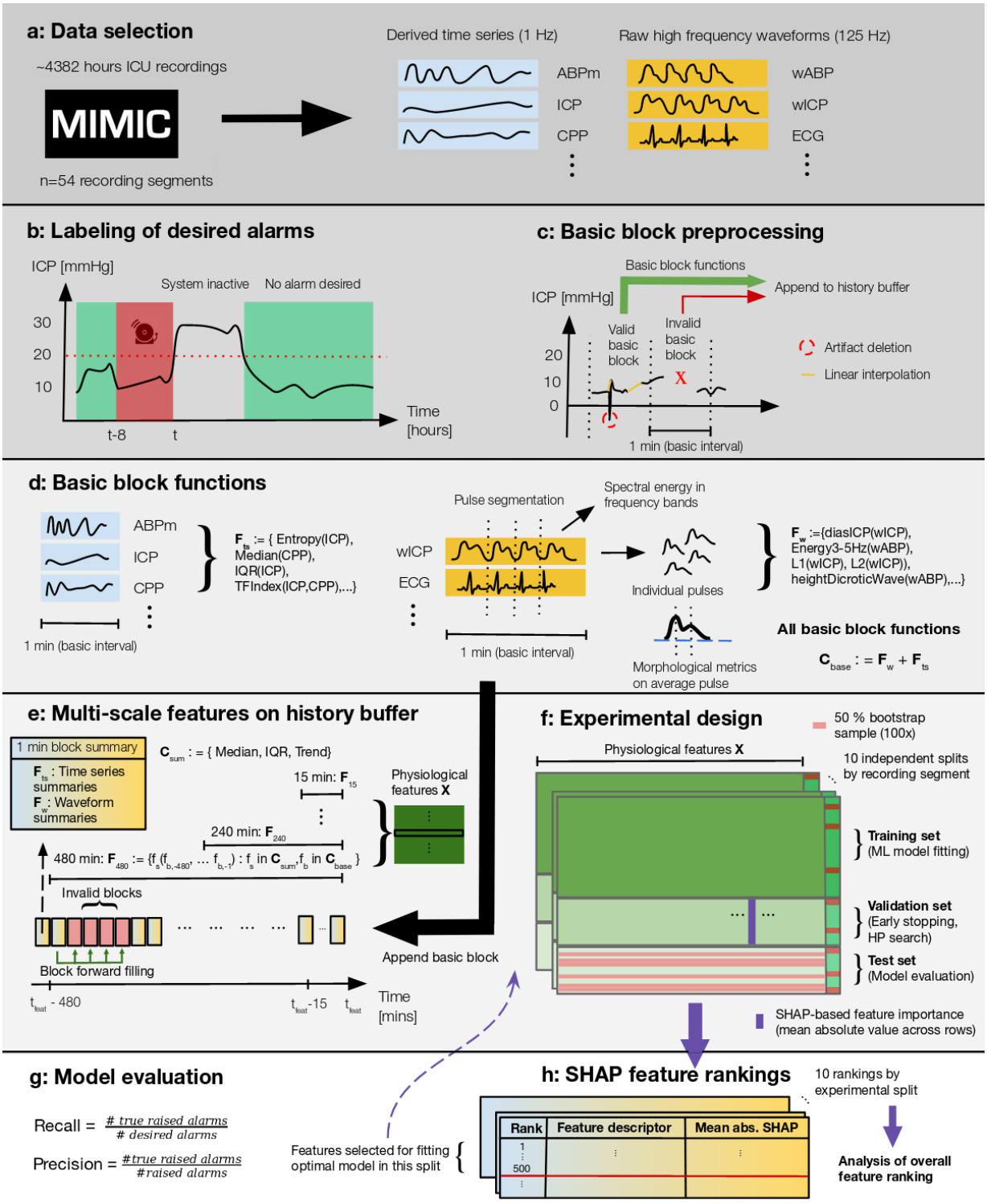

Fig. 3. Overview of physiological feature extraction, alarm labeling and experimental design of the performed study

positive, and 173,194 negative.

D. Physiological feature extraction framework

Basic block functions: During the feature generation process, so-called *basic block functions* are computed online on non-overlapping windows containing 1 minute of high-frequency waveforms/time series, corresponding to 60 samples @1 Hz or 7500 samples @125 Hz (Fig. 3d). The choice of 1 minute as a basic interval makes computation of complex morphological functions tractable, increases robustness to signal artifacts and sensor detachments, and allows to produce updated predictions every minute. Before computing basic block functions, a window is pre-processed by removing physiologically implausible values. If at least half of the

samples are valid, we reconstruct the remaining samples by linear interpolation. Otherwise, invalid basic blocks marked by a symbolic value are emitted (Fig. 3c). Basic block functions are then computed on valid blocks. As basic block functions we have considered statistical/complexity summaries (median, interquartile range(IQR), line length, Shannon entropy), spectral band energies of waveforms, morphological pulse summaries of the wABP and the wICP waveforms, as well as cerebral auto-regulation indices. Morphological pulse metrics are computed by an algorithm consisting of several steps (Fig. 3d). First, individual pulses on wABP/wICP are segmented, using variants of known algorithms [57]–[59], with the help of the ECG waveform as a reference to identify pulse onsets. Valid pulses in the window are then temporally scaled to make their lengths comparable, overlaid and averaged point-wise, yielding an averaged pulse. Morphological pulse metrics, modeled on those described by Hu et al. [20] and Almeida et al. [60], are then computed on the averaged pulse. A complete overview of the basic block functions is provided in Table II.

TABLE II
OVERVIEW OF BASIC BLOCK FUNCTIONS COMPUTED ON 1-MINUTE WINDOWS

Statistical/complexity summaries (ICP, CPP, ABPm/d/s)
Median, Interquartile range, Line length [61], Shannon entropy
Spectral band energy metrics (wICP, wABP, wPLETH, wRESP, ECG)
Energy in frequency bands [0,1],[1,2],[2,3],[3,6],[6,9],[9,12],[12,15] Hz
Autoregulation indices on time series (1 Hz sample rate)
AmpIndex(ICP,ABPm), AmpIndex(ICP,CPP), AmpIndex(CPP,ABPm) [31] PaxIndex(ICP,CPP,ABPm) [31] PrxIndex(ICP,CPP,ABPm) [62, 63] RapIndex(ICP,CPP) [64, 65] SlowWaveIndex(ICP) [66] TFIndex(ICP,ABPm), TFIndex(ICP,CPP), TFIndex(CPP,ABPm) [67]
Autoregulation indices on waveforms (125 Hz sample rate)
AmpIndex(wICP,wABP) [31] SlowWaveIndex(wICP) [66] TFIndex(wICP,wABP) [67] IaacIndex(wICP,wABP) [37]
Morphological pulse metrics on waveforms
wABP pulse descriptor (17 metrics) [60]: A, UpstrokeTime, TimeAtΠ, TimeAtDw, DownstrokeTime, SysDiasTimeDifference, HeightSysPeak, HeightInflPoint, HeightDicroticWave, R1, R2, R3, R4, R5, R6, Aix **wICP pulse descriptor** (20 metrics) [20]: Mean, Dias, DP1, DP2, DP3, DP12, DP13, DP23, L1, L2, L3, L12, L13, L23, Curv1, Curv2, Curv3, Slope, DecayTimeConst, AverageLatency

Multi-scale history: Computed basic block features are appended to a *history buffer* using an online algorithm, with one batch of features appended per minute (Fig. 3e). If a block is invalid or some features cannot be computed, for example due to missing signals, they are forward filled from the last valid feature in the history. If there is no valid feature in the recent past, the feature value is set to the median of that feature value in the accumulated history. After the history buffer is updated, a new sample of machine learning features is emitted by summarizing the current state of the history buffer (Fig. 3e). As summary functions, we use the median (location estimate of a basic feature over the history), IQR (variability of a basic feature over the history) and the slope of regression line fit (trend of a basic feature over the history). These summary functions are applied separately over the last 15, 30, 60, 120, 240, 360 and 480 minutes to capture pre-hypertensive patterns at various scales of the feature buffer history. After the full feature matrix is built, we standardize all feature columns to have zero mean and unit standard deviation, using statistics from the training data-set. Missing values are replaced by zero, which corresponds to global mean imputation. For machine learning models that can deal with missing data natively, like decision trees or tree ensembles, missing data imputation/normalization was not performed. The online signal processing and feature generation algorithms were implemented using the numerical packages SciPy and NumPy in Python 3.6.

E. Feature interpretation using SHAP values

To gain insights into the precursor patterns of intracranial hypertension we have used SHAP value analysis [68] to uncover the most important features that explain ICH predictions. SHAP (SHapley Additive exPlanation values) is a local feature attribution method, which attributes risk scores of future intracranial hypertension to individual signal patterns, encoded in the physiological features. The SHAP value of a feature $x_i = k$ is the expected change of the risk score, when this feature is added to update the risk estimate, integrating over all possible subsets of other variables which are already used in the risk estimate, prior to adding the new variable. The SHAP value of a prediction at $\mathbf{x}$ for feature i is defined as $s_{i,\mathbf{x}} := E_S[E[f(\mathbf{x})|\mathbf{x}_{S\cup\{i\}}] - E[f(\mathbf{x})|\mathbf{x}_S]]$. Here f denotes the risk score, $E[f(x)|\mathbf{x}_S]$ the conditional expectation of the risk score if the values of features in S are fixed to their observed values, and $E_S[\cdot]$ the expectation over the choice of fixed features S. The used TreeSHAP algorithm [69] is an implementation of SHAP values for tree ensembles, which can deal with missing values, and hence simplifies the computations of $s_{i,\mathbf{x}}$. We summarized SHAP values of predictions on the validation set by defining the global importance of a feature as $g_i := n_{\text{val}}^{-1} \sum_{\mathbf{x}\in\text{val}} |s_{i,\mathbf{x}}|$, as the mean magnitude of risk score change that a particular feature causes when introduced into the model (Fig. 3f). Hereby, n_{val} is the number of samples in the validation set. All features were ranked by $\{g_i\}$ for each split, defining the *top features* of the split. Ranks were averaged across splits to increase robustness of the reported feature rankings (Fig. 3g). SHAP values were also used as a *feature selection* method internal to each split, by discarding all but the top 500 features on the validation set before creating derived models (Fig. 3g). In this way, overfitting to non-informative features, which are numerous due to the broad range of feature combinations, is reduced. The test set was not used for feature selection or feature ranking purposes.

F. Machine learning models

As machine learning models we have considered LogReg, a L2-regularized logistic regression model optimized using stochastic gradient descent [70]; Tree, a single decision tree; GradBoost, a gradient-boosted ensemble of decision trees

[71]; and MLP, a multi-layer perceptron with a sigmoid activation function. Implementation details and hyper-parameter search grids for all machine learning models are listed in the supplementary material.

G. Experimental design

Prediction models were evaluated using precision @70, 80 and 90% recalls, which reflects our prior belief that an alarm system for ICH should have high sensitivity, whereas false alarms are more tolerable and can be reduced with post-processing defined on top of the sequence of prediction scores (Fig. 3g). All experimental result tables report these 3 metrics. 95% standard-error-based confidence intervals of performance metrics, which are used in all figures/tables, were constructed by drawing 10 randomized train/validation/test splits (proportion 50:25:25%) with respect to complete recording segments. Splits were stratified, such that the positive label prevalence of training, validation, test sets in each split is within 0.02 of the overall prevalence in the cohort. This minimizes nuisance effects for performance metrics sensitive to label prevalence (e.g. precision). The experiments performed per split are completely independent. The training set was used for model fitting, while the validation set was used for implementing early stopping heuristics, choosing the optimal set of hyperparameters, and computing mean absolute SHAP values that define the reported feature rankings (Fig. 3f). Each split is associated with a distinct feature ranking, which we integrate over in the feature importance analysis (Fig. 3h). The test set was used to compute all reported performance metrics; hyperparameters and optimal features were *not* selected on this set to avoid overfitting. To account for test set variability, besides training process variability, we drew 100 bootstrap samples (size 50% of test-set samples) with replacement from the test set, yielding 1000 overall replicates. Models with (indistinguishable based on overlapping 95% confidence intervals) best performance are listed in bold-face.

IV. Results

Low-frequency time series channels

As a sanity check, we trained several models that do not use any features derived from high-frequency waveforms. The results, shown in the first part of Table III, indicate that ICP is the single most valuable time series across all desired recall levels. The addition of ABP/CPP context information leads to consistent performance increases.

Importance of high-frequency waveform metrics

Taking the most performant time series model (from the first part of Table III), we tested whether adding features derived from 125 Hz waveforms has a positive effect on the prediction performance (second part of Table III). Our results indicate that adding wICP yields a marked performance increase, and the joint use with wABP strengthens this effect even further. Using only waveform channels shows consistently higher performance than just using time series.

TABLE III
Prediction performance of models by inclusion of physiological time series/waveform channels in the feature generation process

Channels	Prec@70Recall	Prec@80Recall	Prec@90Recall
ICP	**0.364** ± 0.007	**0.342** ± 0.005	0.308 ± 0.004
ABP	0.286 ± 0.002	0.283 ± 0.002	0.274 ± 0.002
CPP	0.270 ± 0.002	0.262 ± 0.002	0.259 ± 0.002
ICP+ABP+CPP (1 Hz)	**0.378** ± 0.008	**0.350** ± 0.006	**0.326** ± 0.005
+wICP	0.400 ± 0.005	0.357 ± 0.005	0.315 ± 0.004
+wICP/ABP	**0.442** ± 0.006	**0.405** ± 0.005	**0.347** ± 0.003
+wALL	0.358 ± 0.004	0.329 ± 0.004	0.292 ± 0.003
only wICP	0.395 ± 0.006	0.355 ± 0.005	0.290 ± 0.003
only wICP/wABP	0.428 ± 0.005	0.392 ± 0.003	**0.348** ± 0.004

Morphological, spectral energy metrics and cerebral auto-regulation indices

Morphological pulse metrics, cerebral auto-regulation indices and band energy have each been shown to exhibit characteristic changes prior to hypertensive events. We tested whether such changes can translate into performance benefits when the corresponding features are added to a simple model. Our results are summarized in the first part of Table IV. Incremental additions of feature categories (ordered roughly by computational cost and algorithmic complexity) lead to consistent performance increases across all desired recalls.

TABLE IV
Prediction performance by models based on different basic block functions (first part), and multi-scale history summary functions (second part)

Feature types	Prec@70Recall	Prec@80Recall	Prec@90Recall
Statistical/Complexity	0.364 ± 0.003	0.341 ± 0.003	0.324 ± 0.004
+AutoRegIndices	0.387 ± 0.006	0.365 ± 0.005	0.333 ± 0.004
+BandEnergy	0.404 ± 0.007	0.383 ± 0.006	**0.348** ± 0.005
+PulseMorphology	**0.442** ± 0.006	**0.405** ± 0.005	**0.347** ± 0.003
Location	0.417 ± 0.006	0.387 ± 0.006	0.343 ± 0.005
Loc+Trend	**0.445** ± 0.006	**0.417** ± 0.005	**0.363** ± 0.003
Loc+Trend+Variation	**0.442** ± 0.006	0.405 ± 0.005	0.347 ± 0.003

Multi-scale history summary modes

It has been reported in the literature that variability or trends of individual metrics are predictive of ICH events. Using different history buffer summary functions, we tested whether such features are indeed valuable vs. location estimates. Our results, listed in the second part of Table IV, suggest that trend function provide benefits, whereas additional variability functions have no marginal positive effect.

How much history do we need to store?

Given the benefits of complex waveform features, it is still unclear whether any change in pre-hypertensive patterns occurs during the short- or also long-term history before the event. We tried to answer this question by ablating the set of multi-scale summary functions supported by our framework. Our results (Fig. 4) indicate that there seem to be no clear saturation effects when adding averages/trends over additional length scales, and that the largest performance gains are provided by adding summaries of the last 6 and the last 8 hours of recordings.

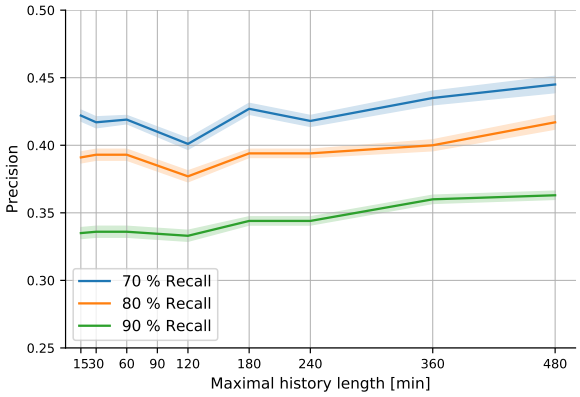

Fig. 4. Model performance conditional on maximum history length contained in model

Comparison of optimal model with baselines

TABLE V
COMPARISON OF DIFFERENT MACHINE LEARNING METHODS APPLIED TO THE OPTIMAL MODEL

ML method	Prec@70Recall	Prec@80Recall	Prec@90Recall
LogReg	**0.445** ± 0.006	**0.417** ± 0.005	**0.363** ± 0.003
MLP	0.421 ± 0.006	0.384 ± 0.005	0.338 ± 0.004
Tree	0.280 ± 0.002	N/A	N/A
GradBoost	0.352 ± 0.002	0.320 ± 0.003	N/A

In a last step, we evaluated different machine learning methods fitted on the optimal model and compared with two baselines from the literature in Tables V and VI. We simulated the method of Hu et al. [20] (**BL1: ICP morphology**) by computing medians of ICP pulse morphology metrics in the last 15 and 30 minutes, which is similar to their pre-hypertensive segment features. A second baseline implements the recently proposed method by Myers et al. [21] (**BL2: Last 2 points + Time to last crisis**) which uses as the three features the last 2 ICP values in a 30 minute window and the time since the last ICH event. If there was no such event, the last feature was set to a large symbolic value. Results show that the simplest machine learning model, i.e. `LogReg`, performed the best, while more complicated models like neural networks (`MLP`) or tree-based methods (`Tree`/`GradBoost`) provided no improvements. Our optimal model significantly outperformed the two baselines, both for PR-based and ROC-based metrics that emphasive high sensitivities. Across all decision thresholds, the AUROCs of the 3 models were 0.736 ± 0.006 (Optimal), 0.712 ± 0.004 (BL1) and 0.718 ± 0.005 (BL2). N/A is shown in Table V if the recall could not be achieved by a machine learning method in all splits.

Ranking of most important physiological metrics

By computing mean absolute SHAP values on the validation set in all 10 splits, we obtained a joint ranking of importance of individual physiological metrics. The 20 most important features for predicting ICH are listed in Table VII. Features that have identical signatures but are computed over distinct scales, are not repeated. Instead, scales that belong to the top 50 features overall are listed in the last column. More details on important features for special feature categories are listed in Tables IX (wICP/wABP waveform) and X (auto-regulation indices).

TABLE VI
COMPARISON OF OUR OPTIMAL MODEL WITH DIFFERENT BASELINES FROM THE LITERATURE (PR-BASED/ROC-BASED METRICS)

	Prec@70Recall	Prec@80Recall	Prec@90Recall
Optimal (LogReg)	**0.445** ± 0.006	**0.417** ± 0.005	**0.363** ± 0.003
BL1: Hu et al. (LogReg) [20]	0.400 ± 0.004	0.362 ± 0.005	0.307 ± 0.004
BL2: Myers et al. (LogReg) [21]	0.402 ± 0.005	0.369 ± 0.005	0.303 ± 0.003
	Spec@70Sens	**Spec@80Sens**	**Spec@90Sens**
as above	**0.673** ± 0.009	**0.590** ± 0.011	**0.449** ± 0.011
	0.634 ± 0.007	0.500 ± 0.010	0.295 ± 0.010
	0.649 ± 0.008	0.544 ± 0.010	0.339 ± 0.008

TABLE VII
OVERVIEW OF 20 MOST IMPORTANT FEATURES FOR PREDICTING ICH

Rank	Feature descriptor	Important scales
1	Med(IcpPulse_Dias(`wICP`))	360,480,240,15,30,120,180
2	Med(SpectralEnergy(`wICP`)_0-1Hz)	480,360,240,180,120,15,30
3	Med(SlowWaveIndex(`wICP`))	480,360
6	Med(IcpPulse_Slope(`wICP`))	480
7	Med(Med(`ICP`))	480,15,180,60
10	Med(IcpPulse_Mean(`wICP`))	480,360
11	Med(SpectralEnergy(`wRESP`)_0-1Hz)	480,360,240
15	Med(SpectralEnergy(`wABP`)_2-3Hz)	480,360
17	Med(TFIndex(`CPP`,`ABPm`))	480
18	Iqr(IcpPulse_Slope(`wICP`))	480
25	Med(SpectralEnergy(`wABP`)_12-15Hz)	360,480
26	Time since admission	N/A
27	Med(SpectralEnergy(`wABP`)_0-1Hz)	480
30	Med(SpectralEnergy(`ECG`)_12-15Hz)	480
31	Iqr(TFIndex(`wICP`,`wABP`))	360
32	Med(SpectralEnergy(`wPLETH`)_0-1Hz)	120
34	Med(SpectralEnergy(`wRESP`)_2-3Hz)	480,360
36	Med(PrxIndex(`ICP`,`CPP`,`ABPm`))	480
39	Current `ICP` median	N/A
42	Iqr(IcpPulse_Dias(`wICP`))	360

As a complementary analysis to feature ablation, we also analyzed which feature categories provided important features according to rankings of mean absolute SHAP values. To enable an easier comparison, we computed the fraction of actual inclusions in the top 200 features (per split) over the number of theoretically possible inclusions. Results are summarized in Table VIII. Waveforms contribute more to highly ranked features than time series, both in absolute and relative terms. In addition, several important features are auto-regulation indices, spectral energies or morphological summaries. Finally, long-scale history summaries between 4 and 8 hours also provide many highly ranked features.

Alarm timeliness before events

The prediction performance of the optimal model for relevant event recalls is shown in Fig. 5. To provide more insights into the behavior of a derived alarm system in clinical settings, we have analyzed the recall of desired alarms before events, conditional on the time until the ICH phase starts. This measures the timeliness of alarms given a fixed model with a constant overall false alarm rate, which is a realistic scenario of clinical implementation. Three false alarm rates that yield sensitive retrieval in excess of 80% across the entire range

TABLE VIII
OVERALL IMPORTANCE OF FEATURE CATEGORIES EVALUATED BY THE NUMBER OF INCLUSIONS IN THE TOP 200 FEATURES, ACROSS 10 SPLITS

Feature descriptor	Inclusion count	Normalized inclusion count
Physiological channel		
wICP	**622**	**0.084**
ECG	**518**	0.077
wABP	**370**	0.057
ICP	193	0.067
wPLETH	166	**0.099**
wRESP	153	**0.091**
CPP	126	0.048
ABPm	98	0.041
ABPs	58	0.060
ABPd	49	0.051
Base feature function		
SpectralEnergy	**819**	**0.098**
IcpPulseMorph	**348**	0.072
AbpPulseMorph	**182**	0.045
Median	172	0.090
Entropy	121	0.063
Iqr	69	0.036
LineLength	61	0.032
TFIndex	60	0.062
SlowWaveIndex	59	**0.123**
AmpIndex	41	0.043
PrxIndex	24	**0.100**
RapIndex	21	0.087
IaacIndex	7	0.029
PaxIndex	0	0.000
Summary function		
Median	**1341**	**0.142**
Iqr	618	0.065
Slope	25	0.003
History length [mins]		
480 Mins	**887**	**0.251**
360 Mins	**534**	**0.151**
240 Mins	**240**	**0.068**
180 Mins	138	0.039
120 Mins	76	0.021
60 Mins	39	0.011
15 Mins	37	0.010
30 Mins	33	0.009

of time-before-events were chosen. As other models did not perform competitively, only the results for LogReg are shown. We can observe a modest decay of alarm recall rates, which stay above 80-90% even 8 hours prior to the event.

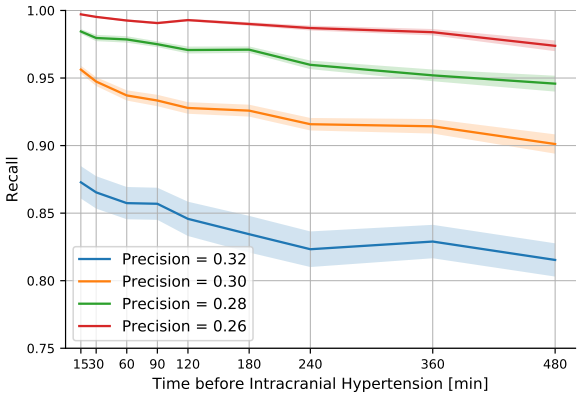

Fig. 5. Evaluation of alarm timeliness before a future ICH event, by time-before-event

TABLE IX
MOST IMPORTANT PHYSIOLOGICAL METRICS EXTRACTED FROM HIGH-FREQUENCY ICP/ABP WAVEFORMS

Rank	Feature descriptor	Most important scales
1	Med(IcpPulse_Dias(wICP))	360,480,240,15,30,120,180
2	Med(SpectralEnergy(wICP)_0-1Hz)	480,360,240,180,120,15,30
6	Med(IcpPulse_Slope(wICP))	480,15,60,30,360
10	Med(IcpPulse_Mean(wICP))	480,360
18	Iqr(IcpPulse_Slope(wICP))	480,360
43	Iqr(IcpPulse_Dias(wICP))	360,480
46	Iqr(IcpPulse_Mean(wICP))	360,480
58	Med(IcpPulse_DecayTimeConst(wICP))	480
79	Iqr(SpectralEnergy(wICP)_0-1Hz)	360
82	Med(SpectralEnergy(wICP)_9-12Hz)	480,360
99	Med(IcpPulse_AverageLatency(wICP))	480
107	Med(IcpPulse_L1(wICP))	480
121	Iqr(SpectralEnergy(wICP)_6-9Hz)	480
137	Med(IcpPulse_DP3(wICP))	480
138	Med(IcpPulse_DP1(wICP))	480
140	Med(IcpPulse_L2(wICP))	480
141	Med(IcpPulse_L3(wICP))	480
144	Med(SpectralEnergy(wICP)_12-15Hz)	480
147	Iqr(IcpPulse_L2(wICP))	360
150	Iqr(IcpPulse_DecayTimeConst(wICP))	480
15	Med(SpectralEnergy(wABP)_2-3Hz)	480,360,240
25	Med(SpectralEnergy(wABP)_12-15Hz)	360,480,240,120
27	Med(SpectralEnergy(wABP)_0-1Hz)	480,360
68	Iqr(AbpPulse_HeightDicroticWave(wABP))	480,360
85	Iqr(AbpPulse_TimeAtII(wABP))	480,360
87	Med(AbpPulse_R4(wABP))	480,360
96	Med(SpectralEnergy(wABP)_9-12Hz)	480,360
103	Med(AbpPulse_HeightInflPoint(wABP))	360,480
120	Iqr(AbpPulse_TimeAtDw(wABP))	360,480
127	Iqr(AbpPulse_R6(wABP))	480
153	Med(AbpPulse_HeightDicroticWave(wABP))	480,360
172	Iqr(SpectralEnergy(wABP)_2-3Hz)	480,360
174	Med(SpectralEnergy(wABP)_1-2Hz)	480
184	Iqr(AbpPulse_R5(wABP))	480
193	Slope(AbpPulse_DownstrokeTime(wABP))	480
194	Iqr(SpectralEnergy(wABP)_9-12Hz)	480
208	Iqr(SpectralEnergy(wABP)_12-15Hz)	480,360
221	Med(AbpPulse_HeightSysPeak(wABP))	360
226	Med(SpectralEnergy(wABP)_3-6Hz)	480
245	Med(AbpPulse_R5(wABP))	480

TABLE X
MOST IMPORTANT CEREBRAL AUTO-REGULATION INDICES EXTRACTED FROM TIME SERIES AND WAVEFORMS

Rank	Feature descriptor	Most important scales
3	Med(SlowWaveIndex(wICP))	480,360,180,240,120
17	Med(TFIndex(CPP,ABPm))	480
31	Iqr(TFIndex(wICP,wABP))	360,480,240
37	Med(PrxIndex(ICP,CPP,ABPm))	480,360,240,180
45	Iqr(SlowWaveIndex(wICP))	480
57	Iqr(AmpIndex(CPP,ABPm))	480,360
67	Iqr(SlowWaveIndex(ICP))	480,360
94	Med(TFIndex(ICP,ABPm))	480,360
106	Med(RapIndex(ICP,CPP))	480,360
111	Med(AmpIndex(CPP,ABPm))	480
129	Med(AmpIndex(ICP,CPP))	480,360
175	Iqr(RapIndex(ICP,CPP))	480,360
204	Med(SlowWaveIndex(ICP))	480
224	Med(AmpIndex(wICP,wABP))	480,360
228	Med(TFIndex(wICP,wABP))	480,360
243	Med(IaacIndex(wICP,wABP))	360,480
294	Iqr(PrxIndex(ICP,CPP,ABPm))	480
309	Med(TFIndex(ICP,CPP))	480
324	Iqr(TFIndex(CPP,ABPm))	480
345	Iqr(TFIndex(ICP,CPP))	480

V. DISCUSSION

We have designed and evaluated a prediction framework for acute intracranial hypertension events, which describes the neurological state using multi-scale descriptors of cerebral autoregulation indices, pulse morphology metrics, spectral energies and statistical summaries. Alarms before critical events were retrieved up to 8 hours before the onset of ICH

(Figure 5) with a recall of 90 % at a precision of 36 %. By analyzing the system using recall/precision we have chosen metrics that more easily translate to the clinical deployment of alarm systems than standard metrics like AUROC. However, to enable easier comparison with previous works we list ROC-based metrics (i.e. relevant thresholds and area under the curve) for the overall results in table VI. The achieved AUROC score for 8 h forecasting is comparable to the work of Myers et al. [21], which is to our knowledge the only published work with a similar forecasting horizon of 6 hours. Yet direct comparison is not easily possible due to substantial difference in cohort size (54 vs. 817 segments) and label definitions. In this work, we implemented the method in [21] to enable a controlled comparison on the same dataset, see Table VI.

Limitations of our study include the inability to match all 54 recording segments to the MIMIC-II clinical database, and hence take clinical context information into account for label and experimental split definition. With respect to labels, by not taking clinical interventions into account, we suspect that we are predicting cases of acute intracranial hypertension that care providers did not anticipate in time before the patient entered the ICH state, which corresponds to scenarios where an early warning system is used as a complementary decision support tool. To analyze this effect in detail, prospectively collected data which documents all clinical interventions and provides contextual information on the ICH events would be required. In our data splits, training, validation and test sets are guaranteed to be temporally disjoint, without interaction between labels and clinical features. Focusing on patients that have comprehensive monitoring with few missing values in the data-set could bias the cohort towards more intensively monitored patients. We chose to apply this criterion regardless because it allows more meaningful conclusions about the utility of different feature channels for predictive modeling.

A novel perspective on the design of ICH alarm systems is provided by the results in Table III. While building a model just using averaged time series defining the event status (`ICP`) provides a good baseline performance, the inclusion of context information (`CPP/ABP`) and richer data modalities like waveforms substantially increased performance. Including high-frequency context information (`wABP`) in addition to `wICP` increased the performance considerably. This hints at new independent information in the ABP waveform and supports the importance of auto-regulation indices, which are functions of two waveforms simultaneously. However, data storage and computational cost associated with waveform data might be considerable, even though our complete framework has faster than real-time performance.

To our knowledge, there is no previous work that assesses the relative merits of different data modalities for ICH prediction. It is an interesting finding that only using waveforms performed better than only averaged time series, especially when compared to recent related works, which found high performance using very simple models, e.g. using only minute-by-minute summaries of ICP.

Each individual feature category among auto-regulation indices, spectral energy, pulse morphology metrics provided marginal performance gains (Table IV) and is relevant for explaining predictions, as assessed using SHAP values (Table VIII). This shows that complex pre-hypertensive patterns previously suspected can translate into relevant machine learning features in our framework. It is surprising that inclusion of variability summaries of basic metrics (Table IV) decreased performance. This might be related to saturation effects and introduction of correlated features, leading to overfitting.

Our results (Figure 4) show a clear trend between the length of considered history and prediction performance, which confirms the design principle of the multi-scale history, and also hints at the relative importance of long-scale physiological changes before hypertensive events, which could inform clinical studies. The same observation can be derived from the analysis of feature category importance in Table VIII, where a clear trend in importance from short-term to long-term features is visible.

The comparison of machine learning models (Table V) provides a pragmatic look at the relevance of the exact statistical learning method for predicting ICH. As has been observed also for other prediction problems in health care, simple models perform surprisingly well, and are not outperformed by models with higher complexity. We suspect that, since the feature choices already incorporate extensive domain knowledge, a simple model like logistic regression is powerful enough. Given the similar performance of `MLP` and `LogReg`, we did not consider the construction of more complex architectures like RNNs or CNNs for this study.

VI. Conclusion

We presented an online machine learning and signal processing framework that forecasts onsets of acute intracranial hypertension up to 8 hours in advance. Using an extensive series of ablation studies we have shed light on the critical components of the framework. SHAP value analysis provided a second perspective on the importance of different feature categories in explaining predictions as well as a ranking of discriminative pattern changes before acute intracranial hypertension. Both perspectives highlight the importance of information derived from waveforms, which provided a substantial performance increase. Our method out-performed two baselines from the literature, which use ICP pulse morphology and 3 simple features of the ICP time series, respectively.

Directions of future work include more sophisticated artifact detection methods at the block level, to minimize the corruption of down-stream feature generation, which is highly sensitive to accurate input signals. Exploring the per-sample SHAP values could provide interpretable reasons for predictions of future ICH events, visualize regions of interest in the history, flag abnormal physiological indices that could precede ICH, as well as generate hypotheses for future studies on the phenomenon. Furthermore, our method could be extended by providing a calibrated alarm system on top of the prediction scores, which triggers alarms at the bedside as a function of sequences of prediction scores. This would be an important step towards clinical implementation of our proposed approach, which was conceived as a real-time algorithm that could directly use data streamed from sensors.

Collecting prospective clinical data to refine the labeling using information on clinical interventions as well as adding clinical co-variates like diagnosis or clinical note concepts could, we suspect, increase the performance of our model even further. Finally, we expect that our framework could be applied to predict other critical events occurring in the injured brain.

ACKNOWLEDGMENT

We acknowledge Emanuela Keller, head of the Neurocritical Care Unit at the University Hospital Zürich, Switzerland, for providing indispensible clinical insights and motivation for this work. MH gratefully acknowledges helpful discussions with Panagiotis Farantatos, Viktor Gal, Stephanie Hyland, Xinrui Lyu, Ngoc M. Pham and Gunnar Rätsch.

REFERENCES

[1] J. A. Langlois *et al.*, "The epidemiology and impact of traumatic brain injury: a brief overview," *J. Head Trauma Rehabil.*, vol. 21, no. 5, pp. 375–378, 2006.

[2] H. M. Bramlett and W. D. Dietrich, "Pathophysiology of cerebral ischemia and brain trauma: similarities and differences," *J. Cereb. Blood Flow Metab.*, vol. 24, no. 2, pp. 133–150, Feb. 2004.

[3] M. Oddo *et al.*, "Brain hypoxia is associated with short-term outcome after severe traumatic brain injury independently of intracranial hypertension and low cerebral perfusion pressure," *Neurosurgery*, vol. 69, no. 5, pp. 1037–1045, Nov. 2011. [Online]. Available: http://dx.doi.org/10.1227/NEU.0b013e3182287ca7

[4] T. Rehman *et al.*, "Rapid progression of traumatic bifrontal contusions to transtentorial herniation: A case report," *Cases J.*, vol. 1, no. 1, p. 203, Oct. 2008. [Online]. Available: http://dx.doi.org/10.1186/1757-1626-1-203

[5] C. Werner and K. Engelhard, "Pathophysiology of traumatic brain injury," *Br. J. Anaesth.*, vol. 99, no. 1, pp. 4–9, Jul. 2007. [Online]. Available: http://dx.doi.org/10.1093/bja/aem131

[6] A. Lavinio and D. K. Menon, "Intracranial pressure: why we monitor it, how to monitor it, what to do with the number and what's the future?" *Current Opinion in Anesthesiology*, vol. 24, no. 2, pp. 117–123, 2011.

[7] N. Carney *et al.*, "Guidelines for the management of severe traumatic brain injury, 4th edition," *Neurosurgery*, vol. 80, no. 1, pp. 6–15, Jan. 2017. [Online]. Available: https://academic.oup.com/neurosurgery/article-abstract/80/1/6/2585042

[8] L. Rangel-Castilla *et al.*, "Cerebral pressure autoregulation in traumatic brain injury," *Neurosurg. Focus*, vol. 25, no. 4, p. E7, Oct. 2008. [Online]. Available: http://dx.doi.org/10.3171/FOC.2008.25.10.E7

[9] L. Rangel-Castillo *et al.*, "Management of intracranial hypertension," *Neurologic clinics*, vol. 26, no. 2, pp. 521–541, 2008.

[10] J. McNames *et al.*, "Precursors to rapid elevations in intracranial pressure," in *Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 4, 2001, pp. 3977–3980. [Online]. Available: http://dx.doi.org/10.1109/IEMBS.2001.1019715

[11] T. F. Bardt *et al.*, "Monitoring of brain tissue PO2 in traumatic brain injury: effect of cerebral hypoxia on outcome," *Acta Neurochir. Suppl.*, vol. 71, pp. 153–156, 1998. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/9779171

[12] S. Badri *et al.*, "Mortality and long-term functional outcome associated with intracranial pressure after traumatic brain injury," *Intensive Care Med.*, vol. 38, no. 11, pp. 1800–1809, 2012.

[13] B. W. Bonds *et al.*, "Predictive value of hyperthermia and intracranial hypertension on neurological outcomes in patients with severe traumatic brain injury," *Brain Inj.*, vol. 29, no. 13-14, pp. 1642–1647, Oct. 2015. [Online]. Available: http://dx.doi.org/10.3109/02699052.2015.1075157

[14] E. Karamanos *et al.*, "Intracranial pressure versus cerebral perfusion pressure as a marker of outcomes in severe head injury: a prospective evaluation," *Am. J. Surg.*, vol. 208, no. 3, pp. 363–371, Sep. 2014. [Online]. Available: http://dx.doi.org/10.1016/j.amjsurg.2013.10.026

[15] M. Majdan *et al.*, "Timing and duration of intracranial hypertension versus outcomes after severe traumatic brain injury," *Minerva Anestesiol.*, vol. 80, no. 12, pp. 1261–1272, Dec. 2014. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/24622160

[16] A. Bhatia and A. K. Gupta, "Neuromonitoring in the intensive care unit. i. intracranial pressure and cerebral blood flow monitoring," *Intensive Care Medicine*, vol. 33, no. 7, pp. 1263–1271, 2007.

[17] R. Sahjpaul and M. Girotti, "Intracranial pressure monitoring in severe traumatic brain injury–results of a canadian survey," *Can. J. Neurol. Sci.*, vol. 27, no. 2, pp. 143–147, 2000.

[18] M.-C. Chambrin *et al.*, "Multicentric study of monitoring alarms in the adult intensive care unit (ICU): a descriptive analysis," *Intensive Care Med.*, vol. 25, no. 12, pp. 1360–1366, Dec. 1999. [Online]. Available: https://doi.org/10.1007/s001340051082

[19] J. McNames *et al.*, "Sensitive precursors to acute episodes of intracranial hypertension," in *The 4th International Workshop in Biosignal Interpretation, Como, Italy*, 2002, pp. 303–306. [Online]. Available: https://pdfs.semanticscholar.org/0cf1/93cd3d50d9882ddd463874b442fc795148ee.pdf

[20] X. Hu *et al.*, "Forecasting ICP elevation based on prescient changes of intracranial pressure waveform morphology," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 5, pp. 1070–1078, 2010.

[21] R. B. Myers *et al.*, "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, 2016.

[22] F. Güiza *et al.*, "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," *Crit. Care Med.*, vol. 41, no. 2, pp. 554–564, 2013.

[23] F. Güiza *et al.*, "Early detection of increased intracranial pressure episodes in traumatic brain injury: External validation in an adult and in a pediatric cohort," *Critical Care Medicine*, vol. 45, no. 3, pp. e316–e320, 2017. [Online]. Available: https://www.ingentaconnect.com/content/wk/ccm/2017/00000045/00000003/art00009

[24] V. De Luca *et al.*, "Temporal prediction of cerebral hypoxia in neurointensive care patients: a feasibility study," in *16th International Symposium on Intracranial Pressure and Neuromonitoring*, 2016.

[25] M. Hüser *et al.*, "Forecasting intracranial hypertension using waveform and time series features," in *Vasospasm 2015 - 13th International Conference on Neurovascular Events after Subarachnoid Hemorrhage*, 2015.

[26] R. Hornero *et al.*, "Interpretation of approximate entropy: analysis of intracranial pressure approximate entropy during acute intracranial hypertension," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 10, pp. 1671–1680, 2005.

[27] J.-Y. Fan *et al.*, "An approach to determining intracranial pressure variability capable of predicting decreased intracranial adaptive capacity in patients with traumatic brain injury," *Biol. Res. Nurs.*, vol. 11, no. 4, pp. 317–324, Apr. 2010. [Online]. Available: http://dx.doi.org/10.1177/1099800409349164

[28] P. Naraei *et al.*, "Toward learning intracranial hypertension through physiological features: A statistical and machine learning approach," in *Intelligent Systems Conference*, 2017, pp. 395–399. [Online]. Available: http://dx.doi.org/10.1109/IntelliSys.2017.8324324

[29] F. A. Zeiler *et al.*, "A description of a new continuous physiological index in traumatic brain injury using the correlation between pulse amplitude of intracranial pressure and cerebral perfusion pressure," *J. Neurotrauma*, vol. 35, no. 7, pp. 963–974, 2018. [Online]. Available: https://www.liebertpub.com/doi/abs/10.1089/neu.2017.5241

[30] M. J. H. Aries *et al.*, "Continuous monitoring of cerebrovascular reactivity using pulse waveform of intracranial pressure," *Neurocrit. Care*, vol. 17, no. 1, pp. 67–76, Aug. 2012. [Online]. Available: http://dx.doi.org/10.1007/s12028-012-9687-z

[31] D. K. Radolovich *et al.*, "Pulsatile intracranial pressure and cerebral autoregulation after traumatic brain injury," *Neurocrit. Care*, vol. 15, no. 3, pp. 379–386, Dec. 2011. [Online]. Available: http://dx.doi.org/10.1007/s12028-011-9553-4

[32] N. Kim *et al.*, "Trending autoregulatory indices during treatment for traumatic brain injury," *J. Clin. Monit. Comput.*, vol. 30, no. 6, pp. 821–831, 2016.

[33] M. Balestreri *et al.*, "Intracranial hypertension: what additional information can be derived from ICP waveform after head injury?" *Acta Neurochir.*, vol. 146, no. 2, pp. 131–141, Feb. 2004. [Online]. Available: http://dx.doi.org/10.1007/s00701-003-0187-y

[34] C. J. Kirkness *et al.*, "Intracranial pressure waveform analysis: clinical and research implications," *J. Neurosci. Nurs.*, vol. 32, no. 5, pp. 271–277, Oct. 2000. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/11089200

[35] F. Scalzo *et al.*, "Intracranial hypertension prediction using extremely randomized decision trees," *Med. Eng. Phys.*, vol. 34, no. 8, pp. 1058–1065, 2012.

[36] R. Hamilton *et al.*, "Forecasting intracranial pressure elevation using pulse waveform morphology," in *Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009, pp. 4331–4334.

[37] P. K. Eide, "A new method for processing of continuous intracranial pressure signals," *Med. Eng. Phys.*, vol. 28, no. 6, pp. 579–587, 2006.

[38] F. Scalzo *et al.*, "Reducing false intracranial pressure alarms using morphological waveform features," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 1, pp. 235–239, Jan. 2013. [Online]. Available: http://dx.doi.org/10.1109/TBME.2012.2210042

[39] F. Scalzo *et al.*, "Intracranial pressure signal morphology: Real-time tracking," *IEEE Pulse*, vol. 3, no. 2, pp. 49–52, March 2012.

[40] X. Hu *et al.*, "Morphological clustering and analysis of continuous intracranial pressure," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 3, pp. 696–705, 2009.

[41] J. Wang *et al.*, "Bag-of-words representation for biomedical time series classification," *Biomed. Signal Process. Control*, vol. 8, no. 6, pp. 634–644, 2013.

[42] P. Xu *et al.*, "Improved wavelet entropy calculation with window functions and its preliminary application to study intracranial pressure," *Comput. Biol. Med.*, vol. 43, no. 5, pp. 425–433, Jun. 2013. [Online]. Available: http://dx.doi.org/10.1016/j.compbiomed.2013.01.022

[43] C.-W. Lu *et al.*, "Complexity of intracranial pressure correlates with outcome after traumatic brain injury," *Brain*, pp. 2399–2408, 2012.

[44] R. M. Jha *et al.*, "Intracranial pressure trajectories: A novel approach to informing severe traumatic brain injury phenotypes," *Crit. Care Med.*, vol. 46, no. 11, pp. 1792–1802, Nov. 2018. [Online]. Available: http://dx.doi.org/10.1097/CCM.0000000000003361

[45] J. Pace *et al.*, "A clinical prediction model for raised intracranial pressure in patients with traumatic brain injuries," *J. Trauma Acute Care Surg.*, vol. 85, no. 2, pp. 380–386, Aug. 2018. [Online]. Available: http://dx.doi.org/10.1097/TA.0000000000001965

[46] D. M. Stein *et al.*, "Use of serum biomarkers to predict secondary insults following severe traumatic brain injury," *Shock*, vol. 37, no. 6, pp. 563–568, Jun. 2012. [Online]. Available: http://dx.doi.org/10.1097/SHK.0b013e3182534f93

[47] A. A. Adamides *et al.*, "Brain tissue lactate elevations predict episodes of intracranial hypertension in patients with traumatic brain injury," *J. Am. Coll. Surg.*, vol. 209, no. 4, pp. 531–539, Oct. 2009. [Online]. Available: http://dx.doi.org/10.1016/j.jamcollsurg.2009.05.028

[48] G. Hergenroeder *et al.*, "Identification of serum biomarkers in brain-injured adults: potential for predicting elevated intracranial pressure," *J. Neurotrauma*, vol. 25, no. 2, pp. 79–93, Feb. 2008. [Online]. Available: http://dx.doi.org/10.1089/neu.2007.0386

[49] B. Quachtran *et al.*, "Detection of intracranial hypertension using deep learning," *Proc. IAPR Int. Conf. Pattern Recogn.*, vol. 2016, pp. 2491–2496, Dec. 2016. [Online]. Available: http://dx.doi.org/10.1109/ICPR.2016.7900010

[50] P. Naraei and A. Sadeghian, "A PCA based feature reduction in intracranial hypertension analysis," in *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*, Apr. 2017, pp. 1–6.

[51] B. W. Bonds *et al.*, "Predicting secondary insults after severe traumatic brain injury," *J. Trauma Acute Care Surg.*, vol. 79, no. 1, pp. 85–90, 2015.

[52] F. Zhang *et al.*, "Artificial neural network based intracranial pressure mean forecast algorithm for medical decision support," in *Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 7111–7114.

[53] J.-S. Shieh *et al.*, "Intracranial pressure model in intensive care unit using a simple recurrent neural network through time," *Neurocomputing*, vol. 57, pp. 239–256, Mar. 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231203005137

[54] F. Zhang *et al.*, "Online ICP forecast for patients with traumatic brain injury," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR)*, Nov. 2012, pp. 37–40. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6460066/

[55] M. Saeed *et al.*, "Multiparameter intelligent monitoring in intensive care II (MIMIC-II): a public-access intensive care unit database," *Crit. Care Med.*, vol. 39, no. 5, pp. 952–952, 2011.

[56] W. C. Ziai *et al.*, "Intracranial hypertension and cerebral perfusion pressure insults in adult hypertensive intraventricular hemorrhage: Occurrence and associations with outcome," *Crit. Care Med.*, vol. 47, no. 8, pp. 1125–1134, Aug. 2019. [Online]. Available: http://dx.doi.org/10.1097/CCM.0000000000003848

[57] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Transactions on Biomedical Engineering*, no. 3, pp. 230–236, 1985.

[58] X. Hu *et al.*, "An algorithm for extracting intracranial pressure latency relative to electrocardiogram R wave," *Physiol. Meas.*, vol. 29, no. 4, pp. 459–459, 2008.

[59] W. Zong *et al.*, "An open-source algorithm to detect onset of arterial blood pressure pulses," in *Computers in Cardiology*, 2003, pp. 259–262.

[60] V. G. Almeida *et al.*, "Machine learning techniques for arterial pressure waveform analysis," *J Pers Med*, vol. 3, no. 2, pp. 82–101, May 2013. [Online]. Available: http://dx.doi.org/10.3390/jpm3020082

[61] R. Esteller *et al.*, "Line length: an efficient feature for seizure onset detection," in *Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2, 2001, pp. 1707–1710.

[62] L. A. Steiner *et al.*, "Continuous monitoring of cerebrovascular pressure reactivity allows determination of optimal cerebral perfusion pressure in patients with traumatic brain injury," *Crit. Care Med.*, vol. 30, no. 4, pp. 733–738, Apr. 2002.

[63] M. Czosnyka *et al.*, "Continuous assessment of the cerebral vasomotor reactivity in head injury," *Neurosurgery*, vol. 41, no. 1, pp. 11–17, Jul. 1997. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/9218290

[64] C. J. Avezaat *et al.*, "Cerebrospinal fluid pulse pressure and intracranial volume-pressure relationships," *J. Neurol. Neurosurg. Psychiatry*, vol. 42, no. 8, pp. 687–700, Aug. 1979. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/490174

[65] D.-J. Kim *et al.*, "Index of cerebrospinal compensatory reserve in hydrocephalus," *Neurosurgery*, vol. 64, no. 3, pp. 494–501, Mar. 2009. [Online]. Available: http://dx.doi.org/10.1227/01.NEU.0000338434.59141.89

[66] J. J. Lemaire *et al.*, "Slow pressure waves in the cranial enclosure," *Acta Neurochir.*, vol. 144, no. 3, pp. 243–254, Mar. 2002. [Online]. Available: http://dx.doi.org/10.1007/s007010200032

[67] R. Zhang *et al.*, "Transfer function analysis of dynamic cerebral autoregulation in humans," *Am. J. Physiol.*, vol. 274, no. 1 Pt 2, pp. H233–241, Jan. 1998. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/9458872

[68] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30*, I. Guyon *et al.*, Eds. Curran Associates, Inc., 2017, pp. 4765–4774. [Online]. Available: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

[69] S. M. Lundberg *et al.*, "Consistent individualized feature attribution for tree ensembles," Feb. 2018. [Online]. Available: http://arxiv.org/abs/1802.03888

[70] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT*. Springer, 2010, pp. 177–186.

[71] G. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems 30*, 2017, pp. 3146–3154. [Online]. Available: http://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree.pdf