GHC: G: Intracranial Pressure Crises: Predicting 30 Minutes Out

Risa B Myers, Advisor: Christopher M Jermaine
Department of Computer Science, Rice University, Houston, Texas

ABSTRACT

Traumatic brain injury (TBI) patients can suffer from episodes of increased intracranial pressure (ICP). Untreated, these crises can result in disability or death. Currently, there is no reliable method of predicting when elevated pressure will occur. Instead, clinicians monitor patients in the intensive care unit and treat based on their judgment and experience.

This paper describes a retrospective study over 817 TBI patients, where the goal is to use 30-minute epochs of patient monitoring data to predict whether or not a patient will enter a period of sustained elevated ICP at 15 minutes to four hours in the future. Methods tested include Gaussian processes, multivariate logistic regression and Bayesian autoregressive modeling. The best model is able to predict the onset of ICP crises with 30 minutes advanced warning with an area under the receiver operating characteristic curve (AUC) of 0.86 using only the intracranial pressure time series and the time since last crisis.

1. PROBLEM AND MOTIVATION

Patients with traumatic brain injuries (TBI) have damage associated with the initial trauma that are often irreversible. However, the chemical and physiological processes initiated by the initial trauma can lead to secondary brain injuries, some of which may be prevented or mitigated. For example, crisis episodes of increased intracranial pressure (ICP) may result in severe brain damage or death. If there were a way to accurately and reliably predict when a patient is going into crisis (a period of sustained elevated ICP), clinicians could be alerted in time to possibly improve patient outcomes.

Predicting elevated periods of ICP is a challenging time series analysis problem. Patients are in crisis only 10-14% of the time, making this a difficult rare class problem. Further, most time series classification methods such as dynamic time warping [2, 6] and shapelets [13] recognize characteristic patterns such as heartbeats or gestures. In the ICP domain, there are no characteristic patterns present in the time series that precede crises. For example, Figure 1 shows the intracranial pressure signal for a TBI patient in the intensive care unit. The crisis that occurs around hour 2.5 is preceded by lower ICP values, followed by a rapid increase. However, the crisis at hour 18 follows a period of volatility centered on the crisis threshold of 20 mmHg.

The goal of this research is to predict future intracranial pressure crises using 30-minute epochs of retrospective patient data. ICP crises are defined as periods of ICP $\geq$ 20 mmHg for at least 15 minutes, as recommended by the Brain Trauma Foundation [3, 11]. This research explores predicting these ICP crises 15 minutes to 4 hours in the future using Gaussian process, logistic regression and a hidden Markov model based algorithm.

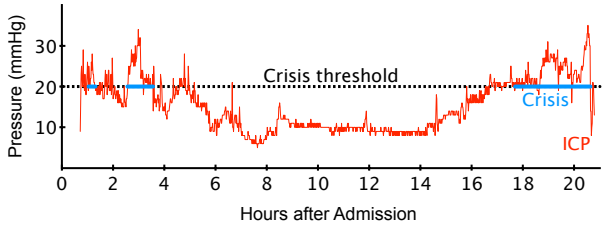


Figure 1: Intensive care unit intracranial pressure (ICP) readings from a traumatic brain injury patient. Crisis periods are represented with blue bars.

2. BACKGROUND AND RELATED WORK

As early as 1983, researchers started collecting ICP values and using them to make predictions. R. Allen [1] advocates short-term prediction of crises to alert clinicians and proposes a number of approaches to detecting significant changes in the ICP signal, including computing the cumulative sum of changes from the mean signal value using a weighted moving average, estimating changes in the mean by applying low-pass filters, such as in the Holt-Winters algorithm or using autoregressive-moving averages. More recent work using a small dataset [10] (30 patients) uses decision trees to predict crises via shape-based features in the signal. This work performs extremely well (AUCs of 0.96, 0.91 and 0.87) but requires data sampled at a very high frequency ($\geq$ 240 Hz) and predicts only 1, 3 and 6 minutes in the future, which is insufficient time to allow for intervention.

The best predictive model for elevated ICP available to date is by Güiza, et al. [4]. These researchers use 4-hour epochs of minute-by-minute ICP and mean arterial pressure (MAP) signals in conjunction with patient length of stay and features derived from the time series (Fourier coefficients, ICP-MAP correlation values, median values over windows of the data, etc.) to predict whether or not a patient will have a crisis 30 minutes after the end of the epoch. Their dataset consists of 239 patients from 22 intensive care units in Europe. The authors are able to predict crises of ICP $\geq$ 30 mmHg for at least 10 minutes, 30 minutes in advance with an AUC of 0.87 using a Gaussian Process model.

3. APPROACH

This section describes the data, features and models used to predict intracranial pressure crises in TBI patients.

3.1 Data Used

The dataset is from 817 traumatic brain injury patients treated at Ben Taub General Hospital in Houston, Texas. For each of these patients, the intracranial pressure values are sampled 100 times per hour. Patients are separated chronologically into training, model selection and evaluation subsets, with 368 training patients, 188 model selection and 261 evaluation. The first two sets were collected from 1989 – 2000 and the evaluation set was collected from 2006 – 2013. This data was segmented into "pre-crisis" or "non-crisis" epochs as shown in Table 1. Changes in treatment protocols and probe technology were minimal over the time period studied.

	Training	Model Selection Cohort	Evaluation Cohort
Total # epochs	43,353	23,751	38,349
# pre-crisis epochs	5,979	3,506	4,025
% pre-crisis epochs	14%	15%	10%

Table 1: Epoch counts for ICP crisis prediction.

3.2 Model Features

Model features came from multiple sources. One set of features was the actual values of the patient monitoring signal and the change in signal in each epoch, i.e. ICP value and delta ICP at time $1, 2, 3, \ldots$ A second set of features was derived from the time series, as defined by Güiza, et al. [4]. These features include median and standard deviation values over segments of the epochs, fast Fourier transform coefficients and frequencies, cepstrum (inverse Fourier transforms) coefficients, and correlation values between the ICP and MAP signals. Finally, features were based on clinical insight, motivated by physiology. One clinical feature used in early experiments was ICP dose [7], that is, the cumulative area under the ICP curve and above the crisis threshold. Another feature was a piecewise linear approximation of "hours since last crisis," as illustrated in Figures 2 and 3, As can be seen in Figure 2, patients who have recently had a crisis are less likely to enter a crisis state, presumably due to the residual impact of the recent therapy. Similarly, patients who have gone a long time without a crisis are, hopefully, improving in health, and are gradually less and less likely to enter another crisis period.

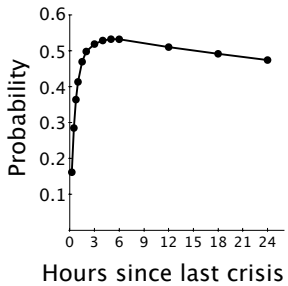


Figure 2: Probability of next crisis.

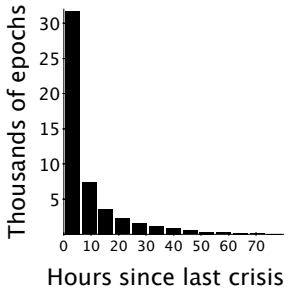


Figure 3: Histogram of hours since last crisis.

3.3 Models

A number of different statistical models (described in detail below) were built and tested on the model selection dataset. The best two models were then used on the final evaluation dataset.

The constructed models include a Gaussian Process (GP) following Güiza's approach, logistic regression and the Autoregressive-Ordinal Regression (AR-OR) model, as described in [8], in combination with logistic regression.

3.3.1 Gaussian Process

The exact same features, covariance functions and process used by Güiza et al. [4, 9] was followed. After evaluating each feature independently using the rational quadratic covariance feature in the gpml toolbox (www.gaussianprocess.org/gpml), the top 5% of the features were used in the final model and the prediction was done using the faster isotropic squared exponential covariance function.

3.3.2 Logistic Regression

The logistic regression model started with the features created for the Gaussian Process experiments, plus the clinical features described in section 3.2. The feature set was reduced by performing feature selection in weka (version 3.7.12 [5]) using information gain and correlation based feature selection.

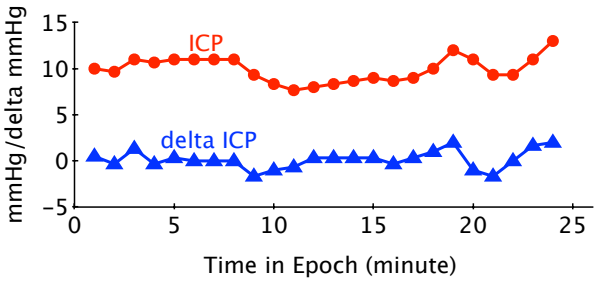
3.3.3 AR-OR model

The AR-OR model is a Bayesian model developed to classify non-pattern based time series [8]. In particular, it was designed to mimic expert assessment of anesthesia vital signs. This model represents time series autoregressively (as the change in value from the previous value) and assigns each time series an ordinal (instead of a binary) label. Labels are determined by fitting a hidden Markov model to the data and learning the state characteristics (mean and variance of emitted values) from the training data. The fractions of time spent in the states, along with other potential features, are used as inputs to an ordinal regression that produces the final labels. In essence, the AR-OR is a dimensionality reduction technique that transforms a multidimensional time series to a small set of numbers, one for the amount of time spent in each state in the model. The number of states is determined empirically.

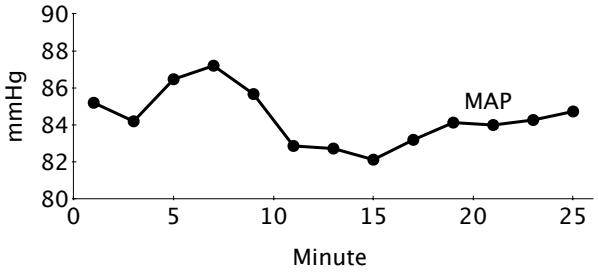
To better predict ICP crises, the AR-OR model was enhanced to use multiple concurrent time series. This functionality allows the use of additional signals, but not to require a direct relationship between them. Figure 4 shows how the ICP and the delta (change in) ICP series are handled as a two-dimensional time series occurring in parallel with a MAP series sampled at half the frequency.

When predicting ICP crises, related patient monitoring signals, including mean arterial pressure central perfusion pressure (CPP) and end tidal carbon dioxide ($EtCO_2$) were used, as these signals are known to be associated with ICP [12].

Assuming that more recent point values are more predictive of the future, a parameter, β, that weights the points in each time series epoch was added, allowing the model to learn how much to emphasize the points later in the series. The weight for each point j is $\beta^{(t-j)}$, where t is the time index $(0, 1, \ldots, 25)$ of the point. The closer β is to 1, the more equally the state assignments are weighted. Figure 5



(a) Intracranial pressure and change in intracranial pressure signals



(b) Mean arterial pressure signal

Figure 4: Concurrent time series

shows how the weight of each point changes as the value of β is changed from 1 (all points being equal) to 0.5. When $\beta = 0.5$, the first point in the epoch only contributes half as much as the last point in the epoch when computing the time spent in each state.

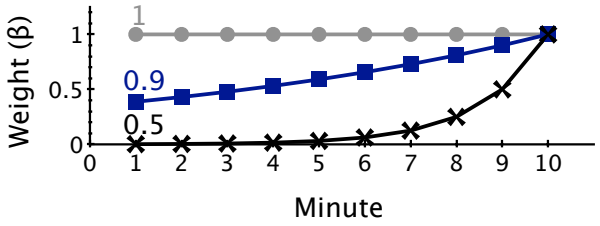


Figure 5: State assignment weights based on the value of β.

3.3.4 AR-OR in Detail

Like all Bayesian models, the AR-OR model relies on a stochastic, generative process that is assumed to have produced the dataset in question. Given the data, the task is then to infer how the process actually produced the dataset. This generative process is embodied by a probability density function (PDF). The probability density function for the AR-OR model has three key components: the model parameters (Θ), representing the states that generate the time series values and the goalposts that separate the different ordinal outcomes; the hidden Markov model state assignments (Y) and the observed ICP values (X) [8].

The PDF for the AR-OR model parameters is:

$$P(\Theta) = \prod_{i=1}^{K} \Big(\text{Dirichlet}(\mathbf{p}_i|.) \times \mathcal{N}(\boldsymbol{\mu}_i|.) \times \text{InvWishart}(\boldsymbol{\Sigma}_i|.) \times \text{Laplace}(r_i|.) \Big) \times \text{Dirichlet}(\mathbf{p}_0|.) \times \prod_{c=1}^{C} \text{Beta}(.)$$

In this equation, K is the total number of states in the model. For each state, $\mathbf{p}_i$ is the transition vector to the other states, $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are the vector of means and the covariance matrix for the multidimensional time series, and r_i is the regression coefficient, which determines how important time spent in state i is to the outcome. Vector $\mathbf{p}_0$ is the probability of the starting states, and C is the number of concurrent time series. The model learns a different exponential weight for each of the concurrent time series, and samples this weight from a Beta distribution. The prior distribution for the transition vectors is a Dirichlet. The vector of mean values for each state's emissions (one for each dimension of the time series) is sampled from a Gaussian distribution and the covariance matrix is initialized from an inverse Wishart distribution. To promote sparsity, the regression coefficients are sampled from a Laplace distribution. The hyperparameters for all of the priors are chosen to be non-informative, as indicated by (.).

The next part of the PDF describes the hidden state values. These assignments reflect which state the patient is in at any given time and are based on the transition matrix that determines the patient's movement from state to state. For example, in a two state model, the patient might be in a non-crisis state or a pre-crisis state, and the non-crisis state might be very stable. That is, a patient in the non-crisis state is unlikely to transition to the pre-crisis state. These parameters are denoted Y. The equation for Y given the model parameters, Θ is:

$$P(Y|\Theta) = \prod_{i=1}^{D} \Big(\text{Categorical}(s_{i,1}|\mathbf{p}_0) \times \prod_{j=2}^{M_i} \text{Categorical}(s_{i,j}|\mathbf{p}_{s_{i,j-1}}) \Big)$$

Here, for each of the D time series, the probability of each state being the initial state is first determined and then each successive state is sampled. M_i is the length of time series i and $s_{i,j}$ is the hidden state assignment for time series i, point j. Note that the state assignments follow the Markov property that the next state assignment is based solely on the previous state, and not on any earlier states values.

Finally, the generative process produces the data for each epoch of time using the hidden state assignments ($\boldsymbol{s}$), the state means ($\boldsymbol{\mu}$) and the covariance matrices ($\boldsymbol{\Sigma}$). The probability of the observed data (X) given the hidden states (Y) and the model parameters (Θ) is:

$$P(X|Y,\Theta) = \prod_{j=1}^{M_i} \mathcal{N}\Big(x_{i,j}|x_{i,j-k} + \boldsymbol{\mu}_{s_{i,j}}, \boldsymbol{\Sigma}_{s_{i,j}} \Big)$$

The AR-OR model is implemented using a Gibbs sampler.

Once the state fractions are computed using the AR-OR model, these values are used in conjunction with the last two ICP points in the epoch and the time since last crisis in a logistic regression model.

The AR-OR models were evaluated two ways, based on the dataset splits of training, model validation and evaluation. The first approach used the model parameters learned from the training dataset exactly. In the second approach, the fractions were learned from the training model and the regression coefficients were refit to the evaluation data using 10-fold cross validation. This approach gives a better idea of how the model will perform when trained on a dataset even more similar to the final evaluation data.

4. RESULTS

Consistent with Allen's recommendation of using short epochs [1], all the predictive models were learned and evaluated using 30-minute epochs of time. To be considered valid, each epoch was required to have at least 25 data points, with no more than 0.05 hours (1.5 minutes) between any two points. Any biologically infeasible values (< 0 or > 80) were considered artifact and were eliminated from the signal. Each epoch was preprocessed by interpolating any missing points to obtain samples every 0.01 hour (the sampling frequency in the provided data), downsampled to every 0.02 hours and smoothed using a first order Savitzky-Golay filter with a window size of 3.

This work was evaluated using the area under the receiver operating characteristic curve (AUC) metric. This measure reflects the ability of a binary classifier to segregate the two outcomes. An AUC of 0.5 reflects guessing the outcome of two equally likely values, while an AUC of 1 represents a perfect classifier.

Figure 6 shows the AUC values for predicting ICP crises over time on the model selection dataset and Figure 7 shows the results of the best two models on the evaluation dataset.

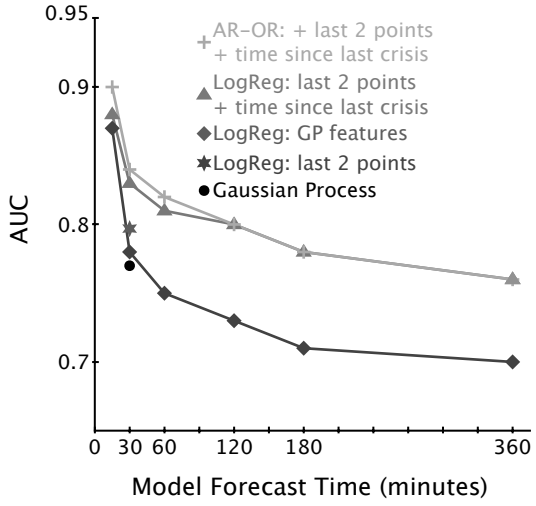


Figure 6: Predictions on model selection dataset.

The Gaussian Process approach was extremely time consuming and produced the worst results on the benchmark 30 minute prediction time (AUC = 0.77). On the other hand, the time since last crisis proved to be a valuable feature, performing better than any individual physiologic signal at prediction times greater than 30 minutes.

Model discrimination was not improved when using signals in addition to ICP.

The best model for predicting ICP crisis events 30 minutes in advance uses a two state AR-OR model to obtain two fractions, along with the time since last crisis and the most recent two ICP values. This model is shown in Figure 8. The arrows in this model are to scale. That is, the heavier the arrow, the more likely that transition will be taken.

While the Gaussian Process experiments did not produce competitive results, as a result of these experiments, they highlighted that the majority of the information associated with the precursors of crises was contained within the last two values of each 30-minute epoch. Consequently, these features were included in other models. Conversely, patient length of stay was excluded, as that metric is controversial as it is frequently driven by factors other than patient health.

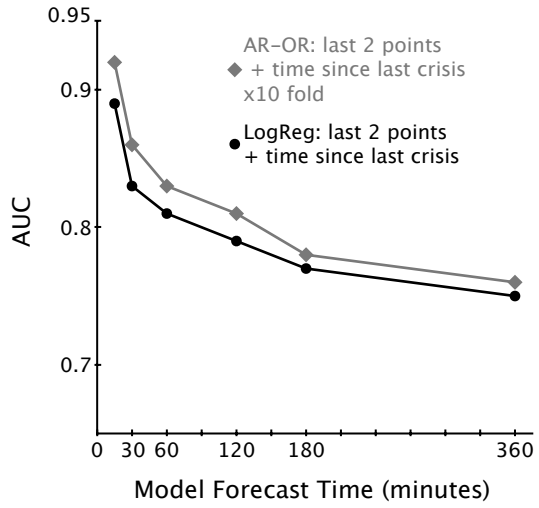


Figure 7: Predictions on evaluation dataset.

As shown in Figure 8, the two states for the ICP model very closely resemble what would be expected – there is a "pre-crisis" state, with mean ICP close to the 20 mmHg crisis threshold and a small positive mean change in ICP, and a "non-crisis" state with mean ICP of 11 mmHg and corresponding very small mean change in ICP.

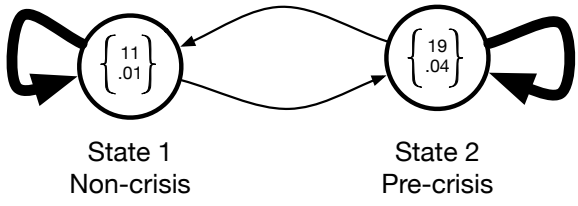


Figure 8: AR-OR state model with time series means (ICP and delta ICP) for predicting intracranial pressure crises.

Figure 9 shows how an evaluation epoch is segmented by state assignment. Based on the time spent in each crisis, the model correctly predicts that the patient is in a crisis 30 minutes later.

5. CONTRIBUTIONS

This work describes a model for classifying pattern-less

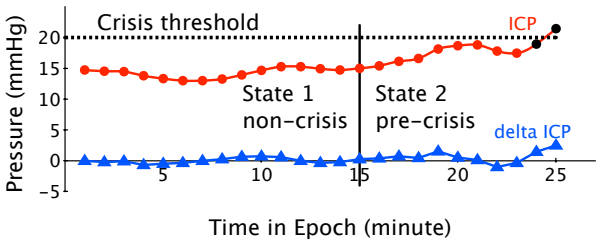


Figure 9: Segmentation of epoch into

time series its application to a rare class classification problem in the medical domain.

5.1 Significance

This work is significant in that it uses very short, 30-minute epochs of time to predict ICP crises 15 minutes to four hours in the future with good performance (AUCs from 0.76 - 0.92). In addition, the results are comprehensible, which is particularly important when working with clinicians. Conversely, the work, by Güiza, et al. starts with a set of over 1,000 potential features, with the end model utilizing 50, including the 68th and the 143rd most recent observation in the 4-hour time series. While these authors' dataset is a good size (239 patients), it is likely that their model is over-fit, as features such as these two specific points are unlikely to generalize well.

5.2 Next Steps

This technique could be applied to other, similarly "patternless" time series, such as weather or financial prediction, in addition to other patient monitoring data. Approaches to handling discontinuities could be explored, so longer epochs can be used. Collaborators at Baylor College of Medicine are hoping to conduct a prospective clinical trial using this predictive model to alert clinicians of potential periods of elevated ICP in TBI patients.

5.3 Conclusions

The enhanced AR-OR model provides a method of reducing time series of arbitrary lengths to a small set of representative values. These values, in conjunction with the last two ICP points and the patient's time since last crisis, comprise a simple and robust predictive model for anticipating sustained periods of elevated intracranial pressure in traumatic brain injury patients. The model is easily understood, accurate and generalizable, as demonstrated by its performance on the final dataset. This model demonstrates the ability of physiological driven data science to produce robust predictive models in the clinical domain.

6. ACKNOWLEDGMENTS

I would like to thank my collaborators at Baylor College of Medicine, without whom this work would not be possible: Craig Rusin, Christos Lazaridus and Claudia Robertson. This research was funded in part by the NSF under grant number 0964526 and by a Gulf Coast Consortia training fellowship on the NLM Training Program in Biomedical Informatics T15LM007093.

7. REFERENCES

[1] R. Allen. Time series methods in the monitoring of intracranial pressure. Part 1: Problems, suggestions for a monitoring scheme and review of appropriate techniques. *Journal of biomedical engineering*, 5(1):5–18, Jan. 1983.

[2] D. Berndt and J. Clifford. Using dynamic time warping to find patterns in time series. *KDD workshop*, 10(16):359–370, 1994.

[3] T. F. Brain et al. Guidelines for the management of severe traumatic brain injury. viii. intracranial pressure thresholds., 2007.

[4] 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: development and validation in a multicenter dataset. *Critical care medicine*, 41(2):554–564, Feb. 2013.

[5] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[6] F. Itakura. Minimum prediction residual principle applied to speech recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 23(1):67–72, 1975.

[7] C. Lazaridis, S. M. Desantis, P. Smielewski, D. K. Menon, P. Hutchinson, J. D. Pickard, and M. Czosnyka. Patient-specific thresholds of intracranial pressure in severe traumatic brain injury. *Journal of neurosurgery*, 120(4):893–900, Apr. 2014.

[8] R. B. Myers, J. C. Frenzel, J. R. Ruiz, and C. M. Jermaine. Correlating Surgical Vital Sign Quality with 30-Day Outcomes using Regression on Time Series Segment Features. In *SIAM International Conference on Data Mining*. Society for Industrial and Applied Mathematics, Apr. 2015.

[9] C. E. Rasmussen and H. Nickisch. Gaussian Processes for Machine Learning (GPML) Toolbox. *The Journal of Machine Learning Research*, 11, Mar. 2010.

[10] F. Scalzo, R. Hamilton, S. Asgari, S. Kim, and X. Hu. Intracranial hypertension prediction using extremely randomized decision trees. *Medical engineering & physics*, 34(8):1058–1065, Oct. 2012.

[11] N. Stocchetti, E. Picetti, M. Berardino, A. Buki, R. M. Chesnut, K. N. Fountas, P. Horn, P. J. Hutchinson, C. Iaccarino, A. G. Kolias, et al. Clinical applications of intracranial pressure monitoring in traumatic brain injury. *Acta neurochirurgica*, 156(8):1615–1622, 2014.

[12] M. Ursino and C. A. Lodi. Interaction among autoregulation, CO2 reactivity, and intracranial pressure: a mathematical model. *The American journal of physiology*, 274(5 Pt 2):H1715–28, May 1998.

[13] L. Ye and E. Keogh. Time series shapelets. In *the 15th ACM SIGKDD international conference*, page 947, New York, New York, USA, 2009. ACM Press.