Permutation entropy analysis of vital signs data for outcome prediction of patients with severe traumatic brain injury[†]

Konstantinos Kalpakis[a*], Shiming Yang[b], Peter F. Hu[b], Colin F. Mackenzie[b], Lynn G. Stansbury[b], Deborah M. Stein[b], Thomas M. Scalea[b]

[a]Department of Computer Science and Electric Engineering, University of Maryland, Baltimore County, MD 21250, [b]University of Maryland School of Medicine, Baltimore, MD 21201

[†]An earlier version of this work has appeared at the 8th International Conference, Machine Learning and Data Mining (MLDM), Berlin, Germany, Jul. 13-20, 2012.

*Corresponding author: Email address: kalpakis@umbc.edu

Abstract:

Permutation entropy is computationally efficient, robust to outliers, and effective to measure complexity of time series. We used this technique to quantify the complexity of continuous vital signs recorded from patients with traumatic brain injury (TBI). Using permutation entropy calculated from early vital signs (initial 10~20% of patient hospital stay time), we built classifiers to predict in-hospital mortality and mobility, measured by 3-month Extended Glasgow Outcome Score (GOSE). Sixty patients with severe TBI produced a skewed dataset that we evaluated for accuracy, sensitivity and specificity. The overall prediction accuracy achieved 91.67% for mortality, and 76.67% for 3-month GOSE in testing datasets, using the leave-one-out cross validation. We also applied Receiver Operating Characteristic analysis to compare classifiers built from different learning methods. Those results support the applicability of permutation entropy in analyzing the dynamic behavior of TBI vital signs for early prediction of mortality and long-term patient outcomes.

1. Introduction

Traumatic brain injury (TBI) is the most common cause of admission to emergency care and trauma-related death in the U.S. civilian population and is a major cause of death and disability in combat causalities [1,2]. In most modern intensive care units (ICUs), vital signs (VS), such as heart rate (HR), blood pressure (BP), and oxygen saturation (SpO2), among others, are collected in high-quality, automated, continuous electronic data streams, as sequential assessments of important physiological functions, providing basic evidence of patients' status. Because VS are an early warning system of physiologic perturbation, they are usually recorded hourly in the ICU setting. However, in most modern ICUs, the massive quantities of high-quality data produced create both a challenge to store, analyze, and interpret and an opportunity to explore novel advanced analytic methods for predicting outcomes. Such predictive algorithms can support advanced instrumentation and decision-assist tools that have the potential to significantly improve clinical outcome for these very ill patients.

To discover the intrinsic patterns that characterize continuous, multivariate, clinical time series, a variety of methods can be used, such as entropy, auto-correlation, autoregressive models, and structure models [3]. One strategy is to embed the time series into higher dimensional space and then compute various entropies for the elements of the embedded time series. Conventional entropies such as Shannon entropy,

Rényi entropy and Tsallis entropy can be calculated given the distribution of elements of the embedded time series. The Rényi entropy of a time series has been used to detect spatially varying multivariate relationships [4] and to study brain injuries [5] and heart rate variability [6,7]. The Tsallis entropy of the elements of a time series has been used to monitor brain injuries after cardiac arrests [8], and to improve the accuracy of gene regulatory networks inference [9].

Bandt and Pompe [10] introduced permutation entropy as a new measure of complexity of non-linear time series. Zanin et al [11] provide an extensive review of various biomedical applications of permutation entropy. Permutation entropy has been used to predict the onset of epileptic episodes from EEG data by considering changes in the permutation entropy of the EEG time series over time [12,13]. Veisi et al [14] find that permutation entropy can be used to effectively classify EEG signals into normal vs. epileptic with an accuracy of 85% even for highly noisy EEG. Physiologically, epileptic episodes/symptoms are manifested with deterministic behavior of the EEG signals, while healthy states are characterized by higher non-chaotic state variability [7]. Permutation entropy has also been used to study sleep using EEGs [15,16], to identify motifs in the EEG signals of patients given fast acting anesthetic drugs [17,18], and to identify temporal gene expression profiles [19]. Bian et al [20] use permutation entropy to identify heart rate variability under different physiological conditions. Berg et al. [21] use ordinal patterns of beat-to-beat heart rate variability from an EKG signals of 40 patients who suffered from myocardial infraction, and try to classify them based on whether they survived for more than two years or not. They achieve a classification accuracy of 85%. Permutation entropy can also highlight forbidden patterns: state-space patterns/permutations appear very infrequently or not at all [22]. It can also be used to quantify non-linear interactions among time series by considering the relative entropy of the joint Takens embeddings of such time series versus the product of independent Takens embeddings.

In this research effort, we use permutation entropy to derive features from continuous, multivariate, time series for outcome prediction of patients with severe TBI. The remainder of this paper is organized as follows. In section 2, we briefly introduce the permutation entropy and the entropy map that we used for quantifying the characteristics of the dynamic system. We use different independence tests to assist in variable selection. In section 3, we describe the dataset and experiment design. We apply the permutation entropy to predict mortality and 3-month Extended Glasgow Outcomes Scale (GOSE), and present experiment results, evaluated by accuracy and the area under the receiver operating characteristic (ROC) curve. We conduct preliminary interpretation of ordinal patterns derived from the VS. Finally, in section 4, we discuss and summarize the results.

2. Method

2.1 Ordinal patterns and permutation entropy

We assume that the physiological status of living things is dynamic but has identifiable and repeated patterns. Likewise, we assume that these patterns will be different in the healthy, injured, and/or ill individuals and that the patterns will be discernibly different from each other. For instance, if the patient is also losing blood, blood pressure (BP) will fall. Heart rate (HR) increases to compensate for the decreased BP to ensure adequate circulation and oxygenation of the brain, and the increase in HR usually increases the BP, at least temporarily. If blood loss continues, BP falls, and clinicians will usually give

fluid, including blood, to raise the BP and ensure adequate oxygenation. These changing patterns of HR and BP are accompanied by changes in intracranial pressure (ICP), cerebral perfusion pressure (CPP), and so on.

Bandt and Pompe [10,23] suggested an approach to time series analysis in which they embedded a continuous time series as a symbolic sequence into another space, a process which they called "permutation entropy". One major ingredient of permutation entropy is the ordinal pattern. The ordinal pattern of a sequence of elements $x_1, \ldots, x_n$ is the permutation (re-arrangement) $\pi = (i_1, i_2, \ldots, i_n)$ that sorts the amplitude values in ascending order so that $x_{i_1} \leq x_{i_2} \leq \cdots \leq x_{i_n}$.

The order L permutation entropy of a time series $x_{1 \ldots N}$ is calculated as follows. Let π_t be the ordinal pattern (i.e. the sorting permutation) for the segment of the time series under the sliding window of length L that ends at x_t, i.e. the subsequence $x_{t-L+1}, \ldots, x_t$. Let $S_L = \{\pi_t\}$ be the set of all those unique (alphabet) ordinal patterns π_t. The time series $x_{1 \ldots N}$ corresponds to the sequence $< \pi_t: t = L, \ldots, N >$ of $N - L + 1$ ordinal patterns from the alphabet S_L. The entropy of this sequence of ordinal patterns is the permutation entropy of the time series $x_{1 \ldots N}$. For example, the Shannon permutation entropy is defined in (1),

$$H_L = - \sum_{\pi_k \in S_L} P(\pi_k) \log\big(P(\pi_k)\big), \tag{1}$$

where $P(\pi_k)$ is the frequency of π_k in the sequence $< \pi_t >$. In the work presented here, we use the Rényi entropy with parameter α of the sequence $< \pi_t >$ defined as

$$R_L^\alpha = \frac{1}{1 - \alpha} \log\left(\sum_{\pi_k \in S_L} P(\pi_k)^\alpha \right). \tag{2}$$

The parameter α in the Rényi entropy acts as a selector of probabilities. It assigns almost equal weight to each possible probability when α is sufficiently close to zero. When α is larger, it puts more weight on higher probabilities. We can use this parameter to assign different weights on events of different probabilities.

2.1.1 Theoretical foundations of permutation entropy

The idea of permutation entropy, introduced by Bandt and Pompe [10,23], relies on a large body of previous work on using information theory to study the phase space (state-space) of dynamical systems [24]. For example, the Kolmogorov-Sinai (KS) entropy is used extensively to characterize the probability distributions (random processes) induced by finite partitions of the state-space of dynamical systems [24,25].

The underlying distribution of the states is an invariant measure of a dynamical system (invariant under smooth transformations of the state space), while entropy functions provide us with a way to compare such distributions. Due to the intractability in deriving explicit analytic expressions of the state distributions, researchers have resorted to numerical estimates from the data. To this end, of particular importance is the Takens-Whitney delay embedding and reconstruction theorem [24] that relates the

dimension d of the system's attractor and the dimension $(2d + 1)$ of the embedding space that is sufficient to reconstruct those properties of the system's attractor that are invariant under smooth transformations. Characterizing the attractor of a dynamical system enables us to predict the system's long-term behavior (since the attractor contains all states that are mapped by the system back into a state in the attractor). The Takens-Whitney theorem provides an effective way to estimate the dimension of the attractor by estimating the Kolmogorov-Sinai entropy of an embedding of a system's state-space.

One particular partition of the space of a Takens delay embedding is obtained via permutations as follows. Consider for simplicity a univariate discrete-time time series x_t, and its Takens delay embedding of order m with delay lag τ: $X_t = \left(x_{t-(m+1)\tau}, .. x_{t-\tau_t}, x_t \right) \in R^m$. Partition the space R^m into m! subsets, each labeled by a unique permutation π of $[1.. m]$, with each subset containing all points in R^m that can be sorted by the subset's labeling permutation.

Permutation-based partitions are more robust to noise and other non-linear distortions and artifacts than value-based fixed-size partitions of the state space, since they depend on the relative order rather than the exact values of the time series. Furthermore, in order to obtain reliable entropy estimates with fixed-size partitions, one needs long time series (in the order of 2^m in order to cover all blocks of such fixed-size partitions); permutation-based entropy estimates do not require long time series. The robustness of permutation entropy makes it particularly attractive for mining vital signs collected in real clinical settings, without expensive pre-processing and cleaning of such signals.

Recall that the uniform distribution has maximum entropy among discrete distributions of bounded support. Large values of permutation entropies correspond to dynamical systems with substantial uncertainty/randomness (divergence in time of initially nearby system states) and small values indicate rather deterministic behavior (fixed points or simple limit cycles).

2.2 Multivatiate time series

Given M variables and a window size M, vital signs within that window are viewed as one slice of size $M \times L$. Fig.1 demonstrates one example of finding ordinal patterns from a finite sequence of time series. Suppose that there are three vital signs ($M = 3$) available for inclusion: ICP, HR, and SBP. Let the window size be $L = 2$. Therefore, one slice constitutes 6 points, which means that we embed VS in a window of size 2 into a higher dimension 6. There are two choices to permute in a slice. The first one considers one slice as one bag. All values in this bag are sorted in ascending order. For example, in Fig.1, slice *a* can be written linearly as the sequence: (ICP)12.4, 14.4; (HR)59.5, 59.4; (SBP)142, 138. Labeling each value 1~6, the values of this sequence can then be sorted into ascending order by applying a permutation $< 1\ 2\ 4\ 3\ 6\ 5 >$. Another choice is to sort within each variable, then concatenate them. For the same example, if we sort ICP, HR, SBP in slice *a* separately, and concatenate their local permutation index, we obtain the pattern $< 1\ 2\ 2\ 1\ 2\ 1 >$. The second method would help keep each variable isolated even if the variables may have similar range, and hence maintain the ordinal patterns from each variable.

2.3 Variable dependencies

More than 50 different types of VS are recorded in the hospital. To make the prediction model simple, we selected a group of VS that are frequently used in clinical diagnosis.

We use tests of statistical independence to explore relations between VS and to select a set of VS for calculating entropy values to sketch the change in patients' physiological status. If two variables have strong linear correlation, then including both variables is not expected to be helpful in capturing more permutation patterns. On the other hand, if two variables are almost independent or have a non-linear relationship, using both may favor including more information.

Recall that two random variables X and Y, with probability density functions $p_X(x)$ and $p_Y(y)$, are independent if and only if their joint density factors as $p_{XY}(x,y) = p_X(x)p_Y(y)$. To evaluate how far away from independence two random variables are, the mutual information (MI) can be used to quantify the difference between $p_{XY}(x,y)$ and $p_X(x)p_Y(y)$ [26]. Recently, Reshef et al.[27] proposed a new measure, the maximal information coefficient (MIC), to assess a wide range of correlations between two variables. MIC takes real values between 0 and 1, representing the two ends of no relationship and noise-free relationship of linear or nonlinear form respectively.

Although MIC is claimed to be powerful to discover a wide range of relationships between two variables, Tibshirani et al.[28] argue that the MIC is inferior to distance correlation [29,30] in many independence tests. In light of this, we compare the MIC and distance correlation coefficients for all pairs of variables.

In Fig.2a, MIC scores are plotted against ρ to show the strength of linear and nonlinear relations. There are some interesting observations that can be made from this plot. First, most of the VS pairs show similar form of relationship despite their mortality outcomes, i.e. the same color dots in the plot cluster in closed regions. Second, most pairs of variables show weak correlations, with MIC and ρ both being close to 0. Third, both MIC and ρ suggest that CPP versus HR and CPP versus SpO2 have relatively stronger linear relationships. We further use the distance correlation to explore the relationships for all pairs of selected vital signs. In Fig.2b, the MIC scores mainly agree with the relationship discovered by the distance correlation coefficients.

Fig.2 indicates that most pairs of variables are weakly dependent or related to each other. Such weak dependency can be further confirmed by the scatter plots of VS pairs. In Fig.3, the scatter plots for some pairs of vital signs show that there are no clear linear nor other functional relationships between those variables. For example, in clinical practice, oxygen is provided essentially constantly to severely injured patients. Most of the time SpO2 stays close to 100 without being strongly related to SBP or other variable. In Fig.3c, we see most points cling to the vertical line SpO2=100%.

2.4 Evaluations

To evaluate results, not only the accuracy, but also the sensitivity, specificity and ROC analysis are utilized to compare performance of different classifiers. The ROC is a tool to depict the tradeoff between sensitivity and specificity. One major reason we adopt the ROC AUC for classifier comparison is that the dataset is skewed, and the ROC AUC is insensitive to the skewness of datasets [31]. This property of ROC curves provides us a way to evaluate the classifiers without worrying about the datasets from which they were trained. Instead of using one single point, we can use the instance statistics to produce a full ROC curve by calculating the class label scores [31]. Provost et al. [32] described a method of calculating the ROC by assigning a score to each instance that reaches the leaf of the decision tree. That score is equal to the ration of positive class labels assigned to that leaf during training. Platt [33] suggested a way

of estimating posterior probability from the output of a support vector machine by fitting a sigmoid function.

3. Experiments and Results

3.1 Data and setup

With the approval from the institutional review board (IRB) of University of Maryland School of Medicine, continuous, automated electronic VS data collected over the course of hospitalization from patients with severe TBI were analyzed using permutation entropy to predict in-hospital mortality and 3-month GOSE outcomes. Our dataset was collected during 2008 and 2009 from 60 sequentially admitted individuals, 9 female and 51 male, 8 of whom died while in hospital. The average duration of stay in hospital was 16 days (ranging from 1.5 to 53 days); 52 patients remained in the hospital longer than 1 week; and 27 patients stayed longer than 2 weeks. Among the 52 patients discharged from the hospital alive, follow-up interviews were carried out at 3 months post-discharge to assess functional outcomes of patients in terms of an 8-category scale [34]: dead, vegetative state, lower severe disability, upper severe disability, lower moderate disability, upper moderate disability, lower good recovery, and upper good recovery. Categories 1 to 4 are defined as "unfavorable" (value 1) and categories 5 to 8 as "favorable" outcomes (value 0). For 3-month GOSE in our dataset, 25 individuals had "favorable" outcome and 35 had "unfavorable" outcome, which, for our purposes, give a relatively balanced dataset.

The raw, every-6-second data were collected by BedMaster® server (Excel Medical Electronics, Jupiter, FL), and were preprocessed to deal with noise due to unstable attachment of sensors, patients' movement and missing values. All data processing and predictive model learning were implemented in Matlab® (2012b, MathWorks, Boston, MA). To reduce the negative effect of noise, VS data were smoothed into a 5-minute tumbling window, as previously [35]. Table 1 shows the percentage of missing points of six selected VS. To utilize all information, we used the k-nearest neighbors' average as surrogate values to fill in missing points.

Determining the optimal selection of VS with which to set up the experiment parameters can be difficult, that is, which values are optimal for the window size and the α range of the Rényi entropy. Therefore, our parameters were selected based on the following considerations. First, a group of VS that are frequently used in clinical diagnosis was chosen, such as ICP, CPP, SBP, SpO2, etc. Those VS with the lowest percentage of missing points and missing data were selected to increase the chances of preserving more patterns, therefore more accurately characterizing the changing physiologic dynamics. A dataset was also tested for change of accuracy with and without removing a given vital sign. Correlated or dependent variables may be included in the dataset for ordinal pattern finding. However, it will not be redundant to include those variables when the relationship among those correlated variables is not order-preserving.

Using the above criteria and tests, a group of five VS were selected (see Table 1) and tested iteratively. The window sizes were equivalent to VS collection duration of 15, of 30 and 60 minutes. In addition, the range for the Rényi entropy parameter α was selected as 0.1 to 2.0 with step size 0.01.

The way to select VS described above is simple and effective but may also raise the question of whether one must include correlated variables in calculating the ordinal patterns. If two variables are not simply linearly correlated, including both may introduce extra ordinal patterns than only using one of them. In

fact, the shock index (SI=HR/SBP) is strongly correlated to HR and SBP, since it is the ratio of the latter two. Because of its nonlinear functional relationship to HR and SBP, adding SI in the ordinal patterns unveils more interesting patterns that portray the interaction between HR and SBP. In general, with the presence of unknown correlation among the variables, the rule of thumb is a straightforward way to help us select a set of VS for study. Using tests of statistical independence, such as MIC or distance correlation coefficient, we also discover and exclude those variable pairs with strong linear relationships. For example, CPP shows relatively strong linear relationships to HR and SpO2 in the independence tests. Also it has very limited contribution to accuracy improvement when it is used with other VS. Hence, CPP was not used in the experiment, although it is an important physiological status indicator.

3.2 Prediction for mortality and 3-month GOSE

With the above setting, experiments were conducted to predict in-hospital mortality and 3-month GOSE. Since the sample size of 60 instances does not form a very large dataset, the leave-one-out cross validation method was used to test the prediction accuracy of our approach to new data.). The leave-one-out cross-validation method is widely used in machine learning studies involving small datasets with 50 to 100 samples [36, pp 203]. Figures 4(b) and 5 (b) show the average ROC curves from the leave-one-out cross-validation.

For each individual patient, a collection of features based on entropy is built as follows. First, selected VS of a certain length (e.g. 3 days VS) [37] are aligned by time and filled in for missing values with the k-nearest neighbor imputation method. Next, given a slice window size L, the VS within a moving window of length L are sorted in bag and are represented by permutations. Such collection of permutations makes an alphabet, where the frequency of each "word" (permutation pattern) is calculated. With a vector of instantiation of parameter α in (2), a set of entropy values is calculated for the window size L. Then the second step is repeated for different parameter values for L. In this manner, a personalized dataset of features is created for each individual patient that encompasses the complexity of the patient's physiological status. With those features, various kinds of classification methods are applied to predict outcomes of clinical interest.

Tables 2a and 2b show confusion matrices and overall accuracy for predicting mortality and 3-month GOSE. The *a priori* knowledge is that 13.3% died in hospital, and 58.3% have unfavorable 3-month GOSE. Using early VS as defined above, a classification tree built upon permutation entropy achieved 62.50% in true positive rate (91.67% in overall accuracy) in predicting death, and 82.86% in true positive rate (76.67% in overall accuracy) in predicting unfavorable cases for 3-month GOSE, which are all higher than the *a priori*. This suggests that the permutation entropy is capable of classifying patients of different physiological status and can handle imbalanced class distribution. On the other hand, the permutation entropy also demonstrates good performance of prediction using early VS. This has potential clinical importance in providing medical care providers with timely prognostic information.

We then applied two other different learning methods, the support vector machine with linear kernel and the quadratic discriminant analysis. The ROC AUC is employed to assess the performance of different classifiers. As noted above, ROC graphs depict the tradeoff between sensitivity and specificity for each classifier in both training and testing datasets, and the AUC measures the probability of the classifier assigning a higher score to the positive than to the negative case, if one positive and one negative case were to be randomly drawn. Figs. 4a and 4b show the in-hospital mortality prediction on the training and

testing sets, using the first 3 days' VS. Figs. 5a and 5b compare prediction power of three classifiers for 3-month GOSE using the last 3 days' VS. Note that the classifier built by the classification tree has the best discrimination for mortality prediction on both the training and the testing sets. The classification tree also has good discrimination capability on the 3-month GOSE outcomes.

3.3 Baseline

In this section, we compare our results with other models created from clinical experience to demonstrate that the permutation entropy method has stable and comparable performance.

Many empirical models have been studied and reported to estimate patients' current and future status. With computer assistance, more statistical metrics can be calculated from long duration vital signs records. Previous work by our group [36,38] on this same dataset studied cumulated dose of ICP >20mmHg, CPP <60mmHg and brain trauma index (BTI=CPP/ICP) as features to predict functional outcomes for patients with severe BTI, using ROC analysis and observed good predictive power for 3-month GOSE 1-4 (AUC=0.65~0.75, p<0.05)[39].

To compare with features built from the permutation entropy, up to five features from the five VS in Table 1 were selected. Mean values of HR, SpO2, SBP, SI, and ICP were calculated using the first 3 days' data for the in-hospital mortality prediction, and the last 3 days for the 3-month GOSE. Table 3 compares the performance of the classification tree built on features from the permutation entropy and the top three classification trees built on subsets of features out of total $\sum_{k=1}^{5} C_5^k = 31$ combinations from Table 1.

It can be observed that the classification tree built upon features created by the permutation entropy demonstrated better performance in terms of overall accuracy and values of AUC for both in-hospital mortality and 3-month GOSE prediction.

3.4 Visualizing frequent patterns

As entropy calculations are not familiar to clinicians, we also consider ordinal patterns with high frequency that are associated with the "dead" outcome.

Since we use five VS, and each vital sign has three points (when using a 15--minute window, with each point corresponding to a 5-minute average), we encode the index of each point of each vital sign in a slice with a two-digit integer 01-15. So, ICP samples have indices 01-06-11, SI samples have indices 02-07-12, SpO2 samples have indices 03-08-13, SBP samples have indices 04-09-14, and HR samples have indices 05-10-15 as shown in Table 4. This encoding allows us to represent the ordinal patterns as a string ("word") of two-digit numbers. Note that an ordinal pattern is a permutation of the 15 two-digit numbers 01-15.

We are interested in the difference of ordinal patterns associated with the "alive" and "dead" outcomes. Let $f_a(P_i)$ and $f_d(P_i)$ be the frequency of pattern P_i associated with the "alive" and "dead" outcome respectively. The ratio $r(P_i) = f_d(P_i)/f_a(P_i)$ measures the degree that P_i is associated with the "dead" outcome. Since the dataset is imbalanced in mortality outcomes, we rescale r by the ratio of the outcome frequencies 52/8. The rescaled ratio is used to rank all the patterns, so that higher ranked patterns are mostly associated with the "dead" outcome. The 10 highest ranked patterns are shown in Table 5.

8

Moreover, we can visualize these ordinal patterns with drawing the changing trends for each vital sign (recall that each ordinal pattern is a permutation of 01--15). Figs.6a~6c, show the top three ordinal patterns as a trend of each vital signs in each 15-minute window. The overall trend of the top 10 patterns is shown in Fig.6d. Such visual patterns together with their VS values may be useful to clinicians for further interpretation.

4. Conclusion

4.1 Summary

Using a large collection of continuous, automated, electronic patient VS data, we derived features to quantify the complexity of this dynamic system using permutation entropy and found that VS features can predict in-hospital mortality and 3-month GOSE, despite a skewed dataset from relatively few instances. These features created by permutation entropy demonstrated promising results. Among 13.3% deaths (58.3% unfavorable cases), we observed 91.67% overall accuracy (62.5% for deaths) for in-hospital mortality prediction, and 76.67% in 3-month GOSE prediction (82.86% for bad outcomes). In comparison with other classifiers on the same dataset, permutation entropy predicted in-hospital mortality and 3-month GOSE with greater accuracy and area under the receiver operating characteristic curves (ROC AUC=0.84, $p<0.001$ for mortality, and ROC AUC=0.71, $p=0.001$ for 3-month GOSE on testing sets).

Permutation entropy is capable of capturing the essentials of dynamic systems described by time series, which can then be used to create interpretable decision rules. Our study demonstrates the capacity of this method to identify, within the first 3 days of care, changes in VS associated with long-term outcome and offers clinicians a potential window for early interventions that may improve patient outcome.

4.2 Future Work

In this study, we used features created by permutation entropy to compare the capabilities of this technique with AUC in prediction of outcome. The accuracy of the prediction models can be improved by including extra descriptive features, such as those features studied in comparison. Furthermore, patients can be categorized into refined subgroups, for which more specific models can be built by categorizing by age or types of injury.

Higher frequency data can be used to enhance early prediction. Optimal calculation of entropy requires time series of sufficient length for a reasonable estimation of ordinal pattern distribution. Using higher frequency data, such as waveform data, permutation entropy may be able to create features to describe the system complexity in earlier time series, such as the first 12 hours in the hospital.

4.3 Clinical Implication

Access to valid clinical prognosis is important in the first 72 hours of care among a group of patients typically hospitalized for several weeks. However, the overall mean time to death for people who died of TBI in our system is 24 hours [2]. Our long-term goal in this work is to provide the critical care team with access to valid clinical prognosis in the first 12 hours after hospital admission and even, if possible, during pre-hospital care and transport, maximizing the potential for timely therapeutic interventions that can save lives and, more importantly, improve long-term clinical outcome.

Reference.

[1] J. J. DuBose, G. Barmparas, K. Inaba, D. M. Stein, T.M. Scalea, L. C. Cancio, J. Cole, B. Eastridge, L. Blackbourne. Isolated severe traumatic brain injury sustained during combat operations: demographics, mortality outcomes and lessons to be learned from contrasts to civilian counterparts. J. of Trauma and Acute Care Surgery, 70 pp. 11-18. 2011

[2] R. P. Dutton, L. G. Stansbury, S. Leone, E. Kramer, J. R. Hess, T. M. Scalea. Trauma mortality in mature trauma systems: are we doing better? An analysis of trauma mortality patterns, 1997-2008. J Trauma 69:(3), pp. 620-626. 2010.

[3] B. D. Fulcher, N. S. Jones. Highly comparative feature-based time-series classification. IEEE Transactions in Knowledge and Data Engineering.pp1-12. 2014.

[4] D. Guo. Local entropy map: A nonparametric approach to detecting spatially varying multivariate relationships. Int. J. Geogr. Inf. Sci., 24, pp. 1367-1389. 2010.

[5] D. Gao, J. Hu, T. Buckley, K. White, C. Hass. Shannon and Renyi entropy to classify effects of mild traumatic brain injury on postural sway. PLoS One, 6:(9). 2011.

[6] Y. Cai, Y. Qiu, L. Wei, W. Zhang, S. Hu, P. R. Smith, V. P. Crabtree, S. Tong, N. V. Thakor, Y. Zhu. Complex character analysis of heart rate variability following brain asphyxia. Med. Eng. Phys, 28:(4), pp. 297-303. 2006.

[7] Y. Cao, W.Tung, J. B. Gao, V. A. Protopopescu, L. M. Hively. Detecting dynamical changes in time series using the permutation entropy. Phys. Rev. E 70:(4). 2004.

[8] A. Bezerianos, S. Tong, N. Thakor. Time-dependent entropy estimation of EEG rhythm changes following brain ischemia. Ann. Biomed Eng., 31:(2), pp. 221-232. 2003.

[9] F. M. Lopes, E. A. de Oliveira, J. R. M. Cesar. Inference of gene regulatory networks from time series by Tsallis entropy. BMC Systems Biology, 5:(61). 2011.

[10] C. Bandt, B. Pompe. Permutation entropy - a natural complexity measure for time series. Phys. Rev. Lett., 88:(17). 2002.

[11] M. Zanin, L. Zunino, O. Rosso, D. Papo. Permutation Entropy and Its Main Biomedical and Econophysics Applications: A Review. Entropy, 14, 1553-1577.2012.

[12] X. Li, G. Ouyang, D. A. Richards. Predictability analysis of absence seizures with permutation entropy. Epilepsy Res., 77:(1) , pp.70-74. 2007.

[13] G. Ouyang, C. Dang, D. A. Richards, X. Li. Ordinal pattern based similarity analysis for EGG reordering. Clin Neurophysiol, 121:(5), pp. 694-703. 2010.

[14] I. Veisi, N. Pariz, A. Karimpour, Fast and Robust Detection of Epilepsy in Noisy EEG Signals Using Permutation Entropy. In Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering, Boston, MA. pp. 200–203, October 2007.

[15] A. A. Bruzzo, B. Gesierich, M. Santi, C. A. Tassinari, N. Birbrumer, G. Rubboli. Permutation entropy to detect vigilance changes and preictal states from scalp EEG in epileptic patients. a preliminary study. Neurol Sci., 29:(1), pp. 3-9. 2008.

[16] N. Nicolaou, J. Georgiou. The use of permutation entropy to characterize sleep electroencephalograms. Clin EEG Neurosci., 42:(1)pp. 24-28. 2011.

[17] X. Li, S. Cui, L. J. Voss. Using permutation entropy to measure the electroencephalographic effects of sevoflurane.Anesthesiology,109:(3), pp. 448-456. 2008.

[18] E. Olofsen, J. W. Sleigh, A. Dahan. Permutation entropy of the electroencephalogram: a measure of anaesthetic drug effect. Br. J. Anaesth, 101:(6), pp. 810-821. 2008.

[19] X. Sun, Y. Zou, V. Nikiforova, J. Kurths, D. Walther. The complexity of gene expression dynamics revealed by permutation entropy. BMC Bioinformatics, 11:607. 2010.

[20] C. Bian, C. Qin, Q.D.Y. Ma, Q. Shen, Modified permutation-entropy analysis of heartbeat dynamics. Phys. Rev. E 2012, 85, 021906:1–021906:7.

[21] S. Berg, S. Luther, S. E. Lehnart, K. Hellenkamp, R. Bauernschmitt, J. Kurths, N. Wessel, U. Parlitz. Comparison of Features Characterizing Beat-to-Beat Time Series. In Proceedings of Biosignal, Berlin, Germany. 2010.

[22] M. Zanin. Forbidden patterns in financial time series. Chaos, 18:(1). 2008.

[23] C. Bandt, Ordinal time series analysis, Ecological Modelling, Vol. 182, No. 3–4, pp 229-238. 2005.

[24] H. Kantz, T. Schreiber, Nonlinear Time Series Analysis. Cambridge University Press, 2003.

[25] Y.G. Sinai, (2007) "Metric Entropy of Dynamical System. http://web.math.princeton.edu/facultypapers/Sinai/MetricEntropy2.pdf

[26] D. J. MacKay. Information Theory, Inference and Learning Algorithms. Cambridge University Press, 2003.

[27] D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, P. C. Sabeti. Detecting novel associations in large data sets, Science 334 pp. 1518-1524. 2011.

[28] N. Simon and R. Tibshirani. Comment on "detecting novel associations in large data sets" by Reshef et al. Science, Dec. 16,, www-stat.stanford.edu/~tibs/reshef/comment.pdf . 2011.

[29] G. J. Szėkely, M. L. Rizzo, N. K. Bakirov. Measuring and testing dependence by correlation of distance. The Annals of Statistics, 35:(6), pp. 2769-2794. 2007.

[30] G. J. Szėkely, M. L. Rizzo. Brownian distance covariance. The Annals of Applied Statistics, 3:(4), pp. 1236-1265. 2009.

[31] T. Fawcelt. ROC graphs: Notes and practical considerations for data mining researchers. Intelligent Enterprise Technologies Laboratory HP Laboratories Palo Alto, HPL-2003-4.

[32] F. Provost, P. Domingos, Well-trained pets: Improving probability estimation trees. 2000.

[33] J. C. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods.In Advances in Large Margin Classifiers, MIT Press, pp. 61-74. 1999.

[34] B. Jennett, J. Snoek, M. R. Bond, N. Brooks. Disability after severe head injury: observations on the use of the glasgow outcome scale. J. Neurol Neurosurg Psychiatry, 44(4), pp. 285--293. 1981.

[35] S. Kahraman, R. P. Dutton, P. Hu, L. Stansbury, et al. Heart rate and pulse pressure variability are associated with intractable intracranial hypertension after severe traumatic brain injury. Clinical investigation, 22:(4).

[36] N. Japkowicz, M. Shah, Evaluating Learning Algorithms: A Classification Perspective. Cambridge University Press, NY, USA, 2011.

[37] D. M. Stein, M. Brenner, P. F. Hu, S. Yang, E. C. Hall, L. G. Stansbury, J. Menaker, T. M. Scalea. Timing of intracranial hypertension following severe traumatic brain injury. J. Neurocritical Care, 18:(3), pp. 332-340. 2013.

[38] S. Kahraman, P. Hu, D. Stein, L. Stansbury, R. Dutton, Y. Xiao, J. Hess, T. Scalea. Dynamic three-dimensional scoring of cerebral perfusion pressure and intracranial pressure provides a brain trauma index that predicts outcome in patients with severe traumatic brain injury. J Trauma, 70:(3), pp. 547-553. 2011.

[39] D. Stein, P. F. Hu, M. Brenner, K. Sheth, et.al. Brief episodes of intracranial hypertension and cerebral hypoperfusion are associated with poor functional outcome after severe traumatic brain injury. Journal of Trauma-Injury Infection & Critical Care, 71:(2), pp. 364-374. 2011.

Table 1: Percentage of available values for selected vital signs

Vital signs	Percentage of available points (%)			
	First 1 day	First 2 days	First 3 days	All
HR	90.07	93.05	94.53	87.60
SpO2	87.04	90.79	92.38	85.20
SBP	88.71	91.80	93.23	81.65
SI=HR/SBP	88.71	91.80	93.23	81.65
ICP	68.63	78.14	79.81	37.72
CPP*	65.69	74.51	76.48	36.45

*not included due to its limited contribution to accuracy.

Table 2: Confusion matrices for classification trees built upon features created by permutation entropy on the testing set.

	First 1 day		First 2 days		First 3 days	
	(A)	(D)	(A)	(D)	(A)	(D)
(A)live	94.23%	5.77%	86.54%	13.46%	96.15%	3.85%
(D)ead	62.50%	37.50%	75.00%	25.00%	37.50%	62.50%
Overall	86.67%		78.33%		**91.67%**	

(a) In-hospital mortality

	Last 3 days		Last 2 days		Last 1 day	
	(G)	(B)	(G)	(B)	(G)	(B)
(G)ood	68.00%	32.00%	44.00%	56.00%	52.00%	48.00%
(B)ad	17.14%	82.86%	37.14%	62.86%	48.57%	51.43%
Overall	**76.67%**		55.00%		51.67%	

(b) 3-month GOSE

Table 3: Comparison between permutation entropy and baseline models on testing set.

Decision tree features	Mortality		
	Accu.(%)	AUC	p-value
Entropy	91.67	0.84	<0.001
ICP/SpO2/HR	85.00	0.71	0.057
SBP	81.67	0.82	<0.001
SI/SBP	80.00	0.78	0.005

Decision tree features	3-month GOSE		
	Accu.(%)	AUC	p-value
Entropy	76.67	0.71	0.001
ICP/SpO2/HR	68.33	0.69	0.005
SBP	65.00	0.68	0.009
SI/SBP	63.33	0.67	0.013

Table 4: Indices of vital signs used in the encoding of frequent ordinal patterns.

	t_1	t_2	t_3
ICP:	01	06	11
SI:	02	07	12
SpO2	03	08	13
SBP	04	09	14
HR	05	10	15

Table 5: Frequent ordinal patterns of the ICP, SI, SpO2, SBP, and HR vital signs associated with the mortality outcome.

P_1	02 07 12 01 06 11 13 08 03 14 09 04 05 10 15
P_2	02 07 12 01 06 11 13 08 03 05 15 10 14 09 04
P_3	12 07 02 01 06 11 08 13 03 05 10 15 04 09 14
P_4	02 07 12 01 06 11 13 08 03 10 05 15 14 09 04
P_5	02 07 12 01 06 11 14 09 04 03 08 13 15 10 05
P_6	12 02 07 01 06 11 08 03 13 15 05 10 09 04 14
P_7	02 07 12 01 06 11 08 03 13 15 10 05 14 09 04
P_8	02 12 07 11 06 01 03 08 13 14 09 04 15 10 05
P_9	07 12 02 11 06 01 03 08 13 04 09 14 10 15 05
P_{10}	12 02 07 01 06 11 09 14 04 03 08 13 15 10 05
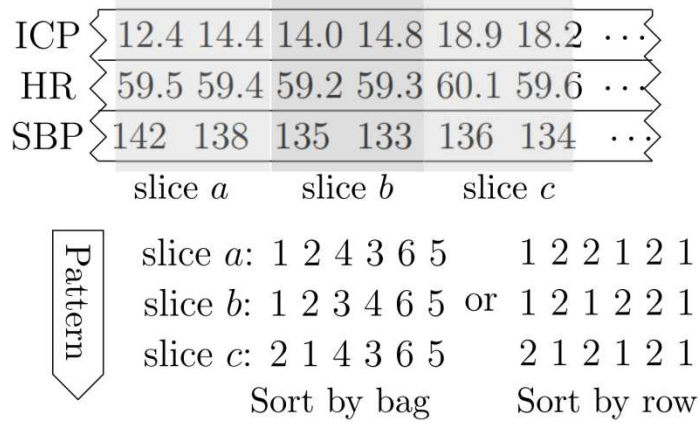
Figure 1: Illustration of ordinal patterns built by permutation in two ways. The exemplary time series snippet comes from 6 points of 5-min smoothed data from one patient.
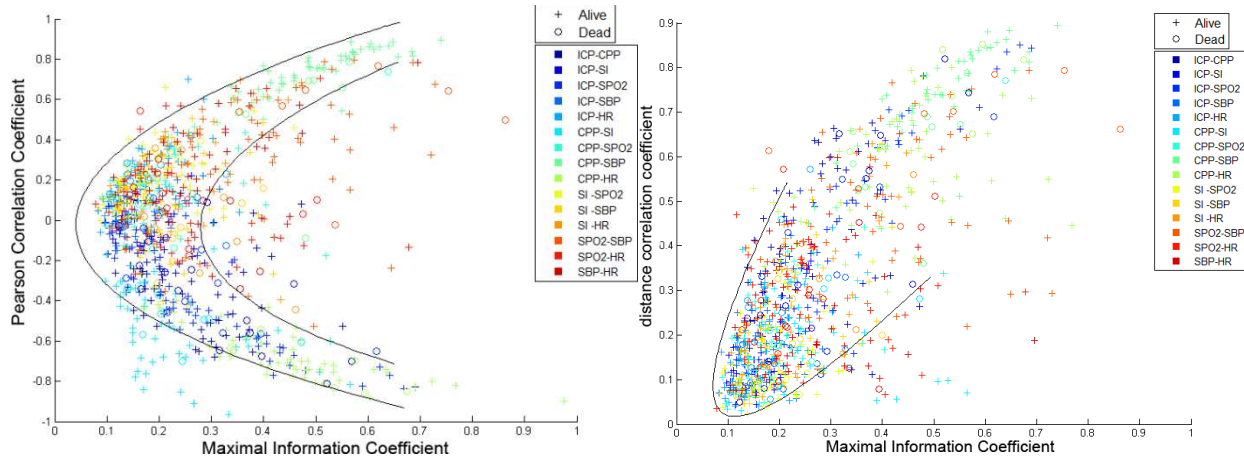
(a) MIC versus ρ (b) MIC versus dCor

Figure 2: Pairwise relationship of six vital signs from each patient, measured by (a) MIC and Pearson's (linear) correlation coefficient ρ, (b) MIC and distance correlation coefficient (dCor).
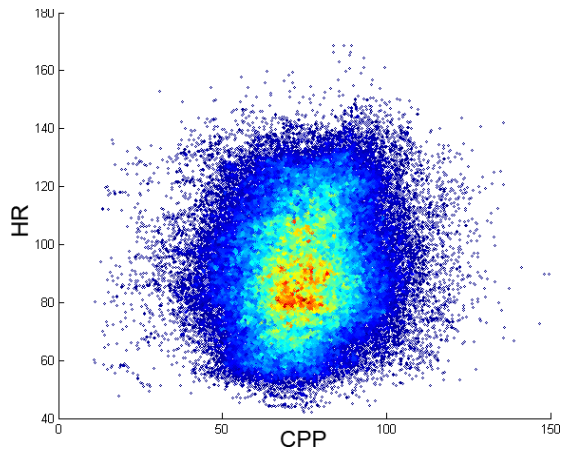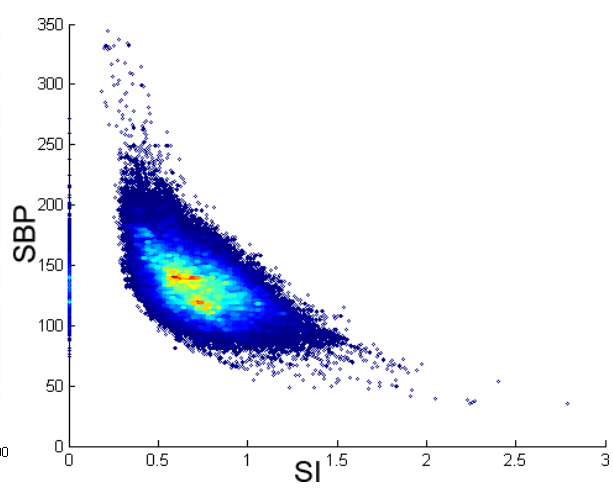
(a) CPP vs HR
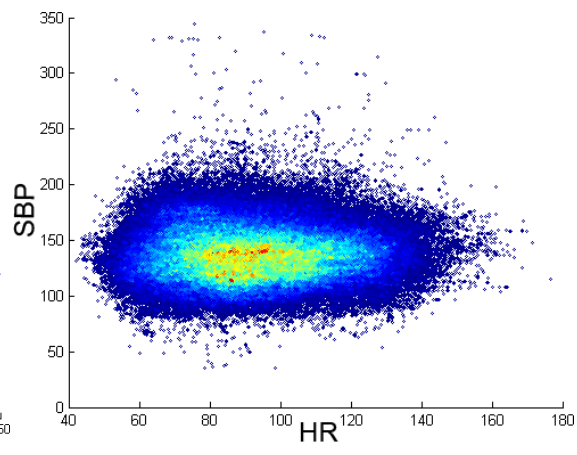
(b) HR vs SBP
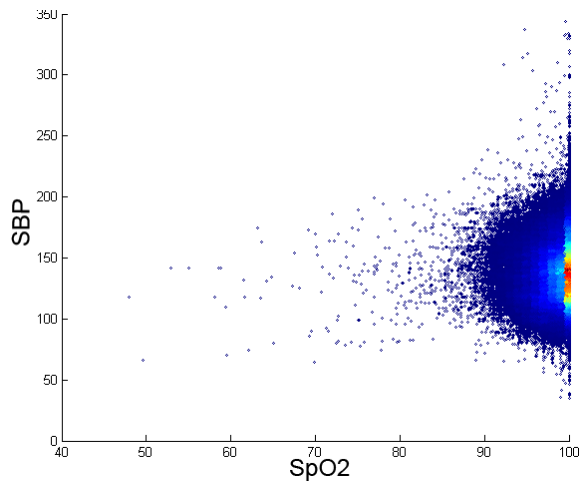
(c) SpO2 vs SBP

(d) SI vs SBP

Figure 3: Scatter plots of different vital signs pairs. Warmer colors imply higher density.
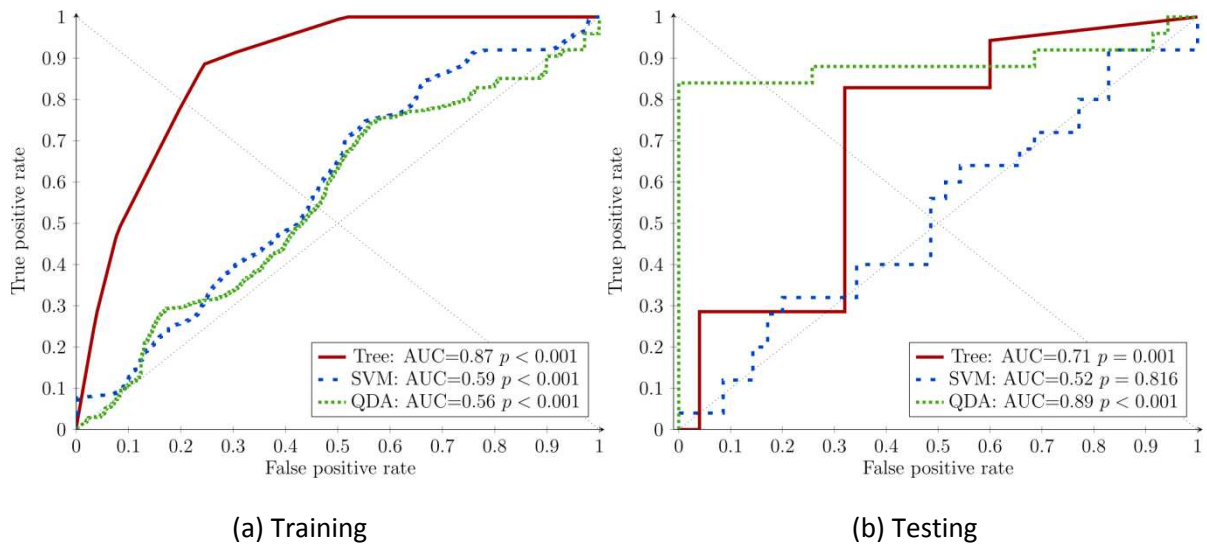
<div align="center">(a) Training   (b) Testing</div>
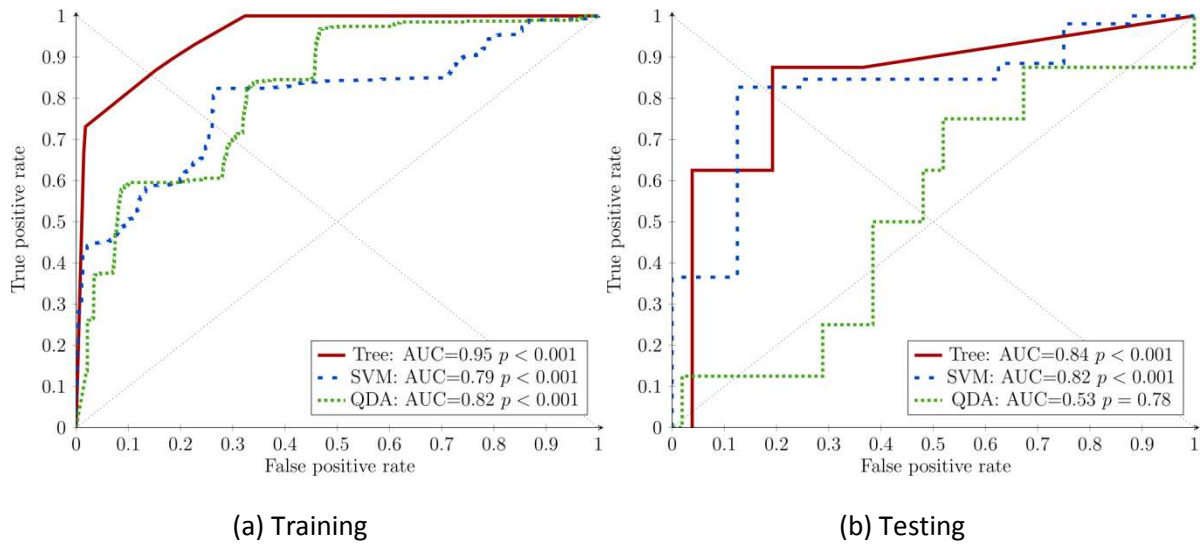
Figure 4: ROCs of mortality classifiers built by three learning methods, using 3 days training set.

<div align="center">(a) Training   (b) Testing</div>

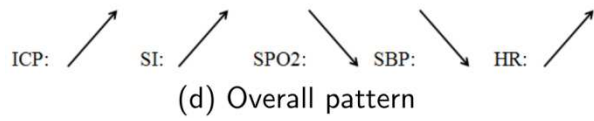Figure 5: ROCs of 3-month GOSE classifiers built by three learning methods, using 3 days training set.

ICP: ↗   SI: ↗   SPO2: ↘   SBP: ↘   HR: ↗

(a) Pattern 1

ICP: ↗   SI: ↗   SPO2: ↘   SBP: ↘   HR: ↗

(b) Pattern 2

ICP: ↗   SI: ↘   SPO2: ↘↗   SBP: ↗   HR: ↗

(c) Pattern 3

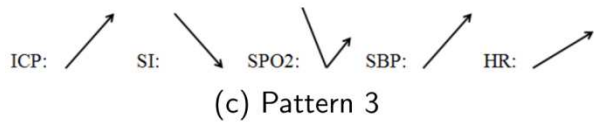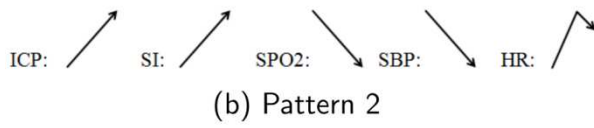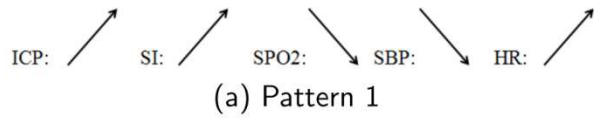ICP: ↗   SI: ↗   SPO2: ↘   SBP: ↘   HR: ↗

(d) Overall pattern

Figure 6: The three highest ranked patterns, mostly associated with the 'dead' outcome, and the overall pattern.