Chapter #: Trauma Outcome Prediction in the Era of Big Data: From Data Collection to Analytics

Authors: Shiming Yang, Peter F. Hu, Colin F. Mackenzie

Abstract:

Physiological data of unprecedented volume is generated daily in healthcare. The Big Data approach in medical data analysis could assist clinicians to do fast and accurate diagnosis, to plan early therapeutic interventions, hence improve patients' outcomes and reduce healthcare cost. However, it is still challenging to reliably collect large amount of data from bedside monitors, and to discover knowledge from the physiological data.

In the area of trauma patient care, early recognition and mitigation of secondary injury and hemorrhage could prevent the death caused by massive bleeding or improve patients' life quality after treatment. We describe methods that are used in a level I regional trauma center for massive physiological data collection and analysis. An almost 100% collection rate of continuous, automated, high fidelity data provides the basis for clinical data mining. Within the machine learning framework, we introduce feature design, feature selection, model performance evaluation, and computing issues, which are critical components in large-scale medical data analysis. Validated knowledge from massive data can be used by clinicians in a simple way for decision making or prioritizing care in the busy hospital environment.

#.1 Introduction

Traumatic brain injury (TBI) and hemorrhage shock are the leading causes of morbidity and mortality after injury both on the battlefield and in civilian care [Holcomb10, Perkin12]. In parallel with new clinical care protocol and technique development, increasing efforts in continuous physiological monitoring and data analysis are reported [Koht12, Reddy15]. The ultimate goal of Big Data's application in this area is to optimize the limited

medical resource allocation through comprehensive analysis of large amount of data, hence improve patients' outcomes and reduce healthcare cost.

Early recognition and mitigation of secondary injury and hemorrhage could ameliorate the effect of brain injury or prevent the death caused by massive bleeding. Various statistical modeling and machine learning methods have been used to explain TBI outcome associations and to predict future conditions with massive datasets. Research and funding agencies, including National Institute of Neurological Disorders and Stroke (NINDS), the U.S. Department of Defense (DoD), are also investing more resources in developing reliable TBI outcome and transfusion prediction models and practically usable tools based on intensive analysis of large data collections. Military medicine considers these approaches as the future way to develop combat casualty autonomous resuscitation [Palmer10,DuBose11] and enhance real-time field decision-making [Provost13].

The volume of real-time physiological patient data has proliferated with each advance in computer hardware and medical sensor technology. High fidelity data are streamed into physiological monitors for care planning, clinical decision-support, quality improvement, and remote patient monitoring. Processing and extracting useful and actionable knowledge from these data also requires consideration of the techniques used to store, manage, and analyze massive data. In this chapter, we will review continuous physiological data collection, signal processing and data analysis methods for predicting trauma patient outcomes related to actionable therapeutic interventions and translating this into autonomous resuscitation. In Section #.2, we describe a reliable large-scale physiological data collection system, which is the basis of any Big Data study. In Section #.3, we review some medical signal processing and feature extraction methods for signals typically used in predicting trauma patients short and long term outcomes. In Section #.4, we provide a description of a comprehensive

framework for machine learning approach to TBI outcomes and blood transfusion prediction, with example projects for the purpose of early detection and automated resuscitation.

#.2 Automated medical big data monitoring

Modern hospitals are often equipped with bedside monitors collecting various physiological data in a real-time, continuous and automated way. Data ranging from routine intermittent observations to high fidelity waveforms can be recorded and streamed into monitors for care planning, clinical decision support [Garg05], quality improvement [Kipnis12], and reduce hospital mortality[Schmidt15]. With massive storage capability, those data can also be stored as part of the electronic health records (EHRs) for retrospective analyses such as physiological pattern discovering [Stein13,Kahraman10b] and prediction modeling [Stein12,Stein11]. One example is the PhysioBank, a large collection of biomedical databases, which inspires studies in cardiovascular time series dynamics, modeling intracranial pressure for noninvasive estimation, and more [Goldberger00, Kashif12, Lehman14, Saeed11]. However, in a busy resuscitation or health-caring environment, collecting more complete data is not the top priority of health providers. System failure or manual data entry errors can result in missing values, causing difficulty in application of decision-support algorithms; or such failures can lose data associated with rare events. Therefore, a reliable system that simplifies and automates the hospital-wise data collection process is necessary.

#.2.1 A real trauma center data

In the R Adams Cowley Shock Trauma Center, a level I regional trauma center located at the downtown Baltimore Maryland, 94 GE-Marquette-Solar-7000/8000® (General Electric, Fairfield, CT) patient vital signs (VS) monitors are networked to provide collection of real time patient VS data streams in 13 trauma resuscitation unit (TRU), 9 operating room (OR), 12 post-anesthesia care (PACU), and 60 intensive care (ICU) individual bed/monitor

units. Each patient monitor collects real-time 240 Hz waveforms and 0.5 Hz trends data which are transferred via secure intranet to a dedicated BedMaster® server (Excel Medical Electronics, Jupiter, FL) and archived [Excel13]. This process generates approximately 20 million data points/day/bed or roughly 30 terabits/year of data. Physiological data collected through this system, when they are displayed on the GE Marquette monitor, include electrocardiographic (ECG), photoplethysmographic (PPG), carbon dioxide (CO_2), arterial blood pressure (ABP), and intracranial pressure (ICP), among others. Trends include heart rate (HR), respiratory rate (RR), temperature, oxygen saturation (SPO_2), end-tidal CO_2 (EtCO2), and ICP, among many others. They cover the categories of brain pressure, cardiac, perfusion, and respiratory.

In addition to continuous data, other heterogeneous data with different formats and temporal resolutions are also collected and organized in databases, including ordinal or categorical data (e.g. Glasgow Coma Scale (GCS), age, sex), radiological images, text (medical records, clinical notes), bed movement (admission, discharge and transfer time and location) and other important data (adverse events, treatments, response, etc.).

#.2.2 Redundant VS collection system with a diagnosis viewer

These heterogeneous data are available and need to be collected with different formats and temporal resolutions. In the current hospital environment, these data are in a loosely organized decentralized network. One approach to manage vital signs waveforms and trend data, used at a Level I trauma centre that admits more than 8000 severely injured patients annually, is to design a triple redundant data collection server for high fault tolerance to maximize these data collection rates to nearly 100%.

To minimize the impact of individual server collection failure, we installed three dedicated BedMaster® servers in parallel to simultaneously collect physiological patient data from the network of patient monitors described above. Figure 1 diagrams the data streams

from multiple individual bed units to three BedMaster® servers arranged in parallel. This triple modular redundancy architecture permits fast switch over time and high system availability [Shooman02]. One server is selected as a principle or "backbone" server. When it fails, values from a second sever will fill in. When two servers fail, values from the third one will be used.

The triple redundant data collection system could increase data availability. However, a tool for fast system diagnosis is still lacking. To address the need for ongoing system status monitoring and real-time presentation critical clinical data, we developed the MoMs (Monitor of Monitors) information representation layer over the VS collection system. Using the current data collecting architecture and a minimum-instrument approach, we stream the most recent record from the BedMaster® server from each bed to a dedicated data server, the MoMs server (Figure 1). A high performance database hosts those data items labeled with data server name, bed unit, timestamp, admission status.

The front-end (MoMs Viewer) is designed as a web-based application so that users can access it from any location in the hospital. IP address white-list and user login modules are used for information security. Each bed collection status is summarized and pushed to the MoMs viewer through the Ajax (asynchronous Javascript and XML) techniques every minute. All participating patient bed units are represented by individual cells in each of 3 spread-sheet blocks representing one of the three redundant BedMaster® servers. Figure 2 shows a block of the web-based monitoring system corresponding to bedside collections from monitors in the TRU, ORs, and neurotrauma critical care (NTCC), and multi-trauma critical care (MTCC) units. The background color of each cell represents the associated bed's data collection status. Green indicates that the data stream has been alive in the last 5 minutes. Yellow indicates that the last timestamp from data from that bed/monitor is 5 minutes to 4 hours old and that a problem may exist. Dark red indicates a timestamp gap greater than 4

hours and that action should be taken to remedy the problem. Report of an elapsed data collection gap includes the duration of collection failure.

There are different elements in each colored cell to indicate bed unit occupancy. Often, nurses may press a bedside button for admission (A) to or discharge (D) from this bed. This allows for a cross-check on potential causes for information gaps such as the device being temporarily inoperable or no patient being monitored. In addition, bed occupancy can be verified by real-time physiological values, such as HR. It can be used as a second evidence for us to infer if a bed unit is currently occupied by a patient. If one bed unit is offline, the gap between now to its last reported time will be shown. Figure 2 shows one example in the unit OR-6, which was highlighted in a red cell with a time gap of 6 hours and 51 minutes.

The easy configuration of the MoMs dashboard viewer also allows it to be used for identifying and displaying clinical information of special research interest. For example, intracranial pressure (ICP) monitoring is an important VS for traumatic brain injured patients and is not often collected due to its invasive nature. To receive early notification of ICP-monitored cases, the MoMs viewer can extract ICP data from all bed/monitor units data streams and display these data using a pre-defined color code. In Figure 2, those pink cells with white bold font text show real-time ICP values from the corresponding bed/monitor units. For example, at the time we viewed the MoMs system, the unit NTCC-17 was monitoring ICP with instant value of 8 mmHg.

In a 12-month study period, single-server collection rates in the 3-month period before MoMs deployment averaged 81.4%. Of the 18.6% collections lost, most (18%) were brief periods, 5 min – 4 hours. Reasons for gaps included collection server failure, software instability, individual bed setting inconsistency, and monitor servicing. In the 6-month post MoMs deployment period, average collection rates were 99.96%. The triple redundant patient data collection system with real-time diagnostic information summarization and

representation improved the reliability of massive clinical data collection to near 100% in a Level I trauma center.

#.3 Data analytics in trauma patients

In treating trauma patients, clinicians often wish to know if a patient has elevated ICP, if some life-saving interventions are needed in the near future, such as massive transfusion, or if a patient have unstable neurologic status. Although high volume of data monitored for trauma patients, they are often underused to answer those questions. Those data are often only used for bedside instant physiological status view. Due to the difficulty to store, access, and analyze, those data streams are hardly used beyond the vital signs readings or instant waveform morphism analysis. Clinicians often utilize those medical data in an empirical way, which may consume their extra attention while not fully unlock the value of the data. Validated, automated medical data processing and analysis could aggregate massive amount of data and assist clinicians to quickly recognize changes in physiological status and prioritize care.

Processing medical data/signals is an important initial step before building predictive models. In pre-hospital and even in hospital, data are collected in hostile environment, which adds noise to the original signals. Numerous methods and algorithms have been studied to remove outliers and noise from the signal, to smooth or sample from high rate signals, and to explore signal characteristics in time, frequency and joint domains [Khawaja06, Kamath12, Drongelen06, Reddy15]. In this section, we assume that data/signals have been pre-processed as 'clean data', and focus on a few topics on feature design, feature selection, and model evaluation, which are less tacked in many medical data processing handbooks.

#.3.1 Features based on descriptive statistics

Maintaining patients' vital signs within normal ranges is a basic task for clinicians. Some treatment protocols also give guidelines on the thresholds of vital signs to be watched

during patient care. For example, for the management of severe traumatic brain injury, it is recommended to initial treatment when ICP is above 20 mm Hg, and CPP is suggested to remain above 70 mm Hg [BTI guide07]. The guidelines for field triage of injured patients recommended to use SBP< 90 mmHg or RR <10 or >29 breaths per minute as a part of the physiological criteria when considering if a patient should be transported to a facility with the highest level of care [Sasser11]. When adding the age factor, SBP < 110 mmHg was recommended in the National Trauma triage protocol for geriatric trauma patients (age > 65 years old) [Sasser11, Brown14]. Empirical thresholds are also used as simple predictors or decision triggers by clinicians. In field triage, SI > 1 is used as an indicator of circulation failure and a predictor of critical bleeding, since it has high specificity and is easy to calculate [Olaussen14].

Those thresholds that pass muster with the clinical experts indeed can serve as domain knowledge to design features from physiological time series. Such type of features may hold clinical meanings that are easy to be interpreted by humans. To quantify the cumulative effect of VS away from normal range, the “pressure times time does” (PTD) is defined as the integrated area enclosed by the VS curve and the threshold line within a given time interval. Sometime, to compare between patients, it is also calculated as averaged PTD in unit time, which is the PTD normalized by time duration. Even thought two patients may be monitored of different time duration, their PTD in unit time still comparable. In predicting TBI patients outcome, the PTD of ICP>20 mmHg and CPP <60 mmHg have been shown to be good predictors of in-hospital mortality and length of ICU stay [Kahraman10].

Similar to the idea of PTD, some descriptive statistics of VS within a given time period are also informative for clinician use or may contribute to outcome prediction in a model. With the assumption that the observed data are approximately normal distributed, mean and variance are often used to sketch the VS value distribution. Standard deviation

(SD) is used to quantify the variability of observed VS data. The coefficient of variance, which is the SD divided by the mean, is a unit-less value that suitable to compare between datasets with widely different means. Robust statistics, such as percentiles or quartiles, are also used to quantify the shape of the VS data distribution. Median (50 percentile or 2nd quartile) is one of the most commonly used statistics in VS feature calculation.

Those descriptive statistics, plus the PTD value of VS in a given time period have the following merits. First, those features can easily incorporate domain knowledge and hence possess straightforward meaning. Stein et al. described a scheme of designing 588 features from 9 typically used VSs, using various thresholds and time windows [Stein12]. Second, they could aggregate large amount of data or simplify complex data as a few summary quantities, which can be used by many statistical prediction models. For example, to identify patients with critical bleeding risk, Mackenzie et al. summarize patients' first 15 minutes 240Hz photoplethysmographic (PPG) by applying above descriptive statistics to the peak-to-valley distances of PPG. In such a way, each patient's 216,000 waveform data points are aggregated into dozens of features as inputs for logistic regression [Mackenzie14]. Third, those values can be calculated efficiently even if the data are of high volume and high velocity.

Despite these listed merits, we need to understand that some of these features rely on certain assumptions. For example, normality is often assumed when we use the mean and SD to sketch the shape of VS data distribution. Besides, we may assume the thresholds suggested by guidelines can be applied to any patient group at any time period. In reality, the thresholds may vary along time, no matter how slow. Also, some hidden factors may result in large difference of threshold values between age groups, or sex groups. Therefore, we need to be careful when incorporating clinical prior knowledge into those features, by examining the applicable conditions.

#.3.2 Features based on variability

Continuous non-invasive ECG and PPG sensors are ubiquitous in both prehospital emergency medical service and in-hospital health care for TBI patients. Waveform measured from both sensors capture rich information of cardiovascular, circulatory and respiratory systems. Heart rate variability (HRV), derived from ECG waveform, has long been used in studies of prehospital life saving interventions [Liu14], neurologic disorder [Ernst14, Acharya06] and the association between the autonomic nervous system (ANS) and cardiovascular mortality is well established [TaskForceESC96]. ANS is a complex life sustaining system which plays a role in nearly every organ and disease [Aslanidia15], including regulating cardiac activity, respiration and pupillary response. ANS dysfunction is a potential complication following severe TBI [Goodman13]. While neurological diseases can lead to many changes in cardiac function, the major changes noted are arrhythmias and repolarization changes. Goldstein et al. [Goldstein96] noted that either increased ICP or decreased CPP can be associated with ANS dysfunction. Bagley et al. [Baguley06] observed that in TBI, patients with and without dysautonomia showed HRV differences compared to control group.

Another universally used sensor, the pulse oximetry, generates PPG waveform that carries rich physiological information, such as heart rate, oxygen saturation, and even respiratory rate [Allen07]. The peaks in PPG are almost synchronized and corresponding to the R peaks from ECG in time domain. Therefore, the peak-peak interval from PPG can be used as an alternative to the NN interval calculated from ECG recordings. Lu et al compared HRV and PPG variability (PPGV) in both time and frequency domain based on data collected from 20 healthy subjects. They found that PPGV were highly correlated to HRV and could serve as an alternative measurement [Lu08].

To find QRS-peaks from ECG, the Pan Tompkins method is used [Pan85]. The ECG signal is passed through a low-pass and high-pass filter to remove noise. The R-peaks are then detected based on a threshold that is adaptive to the signal. To detect PPG signal peaks and valleys, the Savitzky-Golay filter is applied to smooth the signal [Press07]. The peaks can be found through two rounds of searching. In the first round, the peaks are roughly found through Matlab build-in routine 'findpeaks'. The median distance between two consecutive peaks PD_{median} is then calculated. In the second round, any small peaks within a range of $0.6\times PD_{median}$ from a large peak would be ignored. Figure 3 shows a typical PQRST segment from ECG, with identified peaks and five items that we use for calculation. The normal-to-normal (NN) interval illustrated by item 1 is the time interval between two consecutive R peaks. HRV variables in time domain and nonlinear dynamics can be calculated based on the Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology [TaskForceESC96]. From items 2 and 4 in Figure 3, the rising time from Q to R and the falling time from R to S can be calculated. Similarly, from items 3 and 5, the rising and falling amplitudes from Q to R and R to S can be calculated.

Because signals may be collected when the patient went through resuscitation, or had significant movement, artifacts exist and do not reflect the patient real physiological status. To flag out signals with large amount of artifact, signal quality can be evaluated based on R-peaks in ECG and the peaks in PPG. The assumption is that signal quality signals have normal distributed R-R intervals. R-R intervals from segments of low quality are detected as outliers using the Z-test. Figure 4 illustrates two segments of signals. The left side shows ECG and PPG with low signal quality flagged with red horizontal bars. The right side shows precisely identified peaks in both signals.

PPGV and waveform morphology feature can also be designed similarly with

expansion based on PPG unique characteristics. Figure 5 (top subplot) shows a normal PPG segment with identified peaks and valleys. Item 1 illustrates a peak-to-peak time interval, which is analog to the NN interval in ECG. PPGV variables and morphology features were calculated from items 1-4 as we did for ECG. PPG waveform also has unique dicrotic notch, and its shape has been studied and shown to be related to arterial stiffness and aging [Voss15]. To measure the deceleration and acceleration near the dicrotic notch, the first and second derivatives of PPG were calculated through three-point central difference [Yousef12, Elgendi14]. The item 6 in Figure 5 (shadowed area) depicts the region where PPG height changes its moving direction with reducing acceleration and speed until it reaches to a local peak. The duration and amplitude change in this movement were derived as two features [Nitzan97].

Similarly, blood pressure variability (BPV) has been derived from noninvasive blood pressure (NIBP) to as cardiovascular risk factors [Voss12]. ABP waveform also carries pulse wave. With identified normal to normal intervals, NIBP or even ABP could be used to calculate BPV in time domain and frequency domain, as well as the nonlinear dynamics features.

#.3.3 Features based on correlations

Components in a biological system interact to each other in a sophisticated way. Discovering patterns that sketch those interactions could help to understand and forecast the biological system behavior. Correlation is a type of statistical quantity that can indicate predictive relationship, linear or functional, bivariate or multivariate. Hence it is a nice tool to explore system component interactions. In this section, we review the applications of different correlations in quantifying physiological system behaviors.

Maintaining normal cerebral perfusion and oxygenation is important in managing severe TBI patients. Monitoring the cerebral autoregulation could inform clinicians if a

patient lost the ability to maintain a constant perfusion when blood pressure changes [Aaslid89]. Cerebrovascular pressure-reactivity index (PRx) was proposed as an indicator of loss of autoregulatory reserve [Czosnyka96, Balestreri04]. PRx is calculated as a moving correlation coefficient between the mean arterial pressure (MAP) and ICP. Given a short time window, about 40 consecutive data points of MAP and ICP in 4~5 minutes are used for calculation [Czosnyka14]. When cerebral autoregulation is intact, CBF remains a normal constant speed and does not change significantly with mean blood pressure. In such situation, PRx should be close to zero, indicating none or weak linear correlation between MAP and ICP. When cerebral autoregulation is damaged after severe head injury, CBF increases or decreases with blood pressure. The absolute value of PRx moves away from zero, indicating strong linear correlation between MAP and ICP. In this way, PRx can serve to continuously monitor the existence of cerebral autoregulation. When indicators of autoregulation are plotted against cerebral perfusion pressure (CPP), a U-shaped curve is generated, consistent with a loss of autoregulation in conditions of hypoperfusion or hyperemia. A study in 2002 by Steiner et al. took advantage of this relationship to construct curves of CPP against PRx, and hypothesized that the minima of these would indicate an ideal CPP at which pressure reactivity is maximized [Steiner02]. This group and others since have validated this model by finding a correlation between patients' deviation away from this optimal CPP and eventual neurologic outcome.

Another example of using linear correlation in brain trauma study is the calculation of pressure-volume compensatory reserve index, called RAP. It is the moving linear correlation coefficient (R) between the amplitude (A) of a frequency component corresponding to heart rate in the ICP waveform (Figure 6) and the mean ICP pressure (P). Given a short time window of length 4~5 minutes, ICP waveform is transformed into frequency domain by Fourier transform. The amplitude of the frequency component relating to the heart rate can be

found. RAP indicates the relationship between ICP and changes in volume of the intracranial space [Balestreri04]. When ICP is low (e.g. IPC<20 mmHg), near zero RAP means a change in cerebral blood volume has no or small impact on ICP, which indicates a good pressure-volume compensatory reserve. After severe head injury, AMP and ICP may be negatively correlated, indicating the loss of cerebral autoregulation.

#.3.4 Features based on entropy

Physiological data are a succession of values outputted from the information source, namely the patient. Entropy is a measure of the amount of information from the source. Given a discrete random variable X with its all possible element set S_X and probability function $p(x)$, the Shannon entropy $H(X)$ of the random variable X is defined as

$$H(X) = -\sum_{x \in S_X} p(x) \log p(x). \quad (\#.1)$$

Given two random variables X and Y, the joint entropy is a generalization to measure the uncertainty of their joint distribution, which is defined as

$$H(X,Y) = -\sum_{x \in S_X} \sum_{y \in S_Y} p(x,y) \log p(x,y). \quad (\#.2)$$

It can also be extended to the joint distribution of a set of more than two random variables. The conditional entropy of random variable X given Y measures the uncertainty of x when y is known, which is defined as

$$H(X|Y) = -\sum_{x \in S_X} \sum_{y \in S_Y} p(x,y) \log p(x|y). \quad (\#.3)$$

There are more extensions derived from above definitions, such as relative entropy, mutual information, conditional mutual information, etc. [Cover12]. They are could serve as useful mathematical tools to extract features from one or jointly more medical time series to quantify the uncertainty and interactions of the systems that they represent.

Approximate entropy (ApEn) and sample entropy (SampEn) are two widely used

entropies in measuring similarity in time series, especially in HRV [richman00,Lake02]. For a physiological time series, even short ones, approximate entropy can be used to estimate the rate of new information it generates [pincus91,richman00]. Given a time series $x_1, x_2 \ldots x_n$, its Taken's embedding vectors with dimension m are $\boldsymbol{X}_m(i) = (x_i, x_{i+1}, \ldots, x_{i+m-1})$, where $i = 1, \ldots, n-m+1$ [Takens81]. The distance $d(\boldsymbol{X}_m(i), \boldsymbol{X}_m(j))$ between any two such points $\boldsymbol{X}_m(i)$ and $\boldsymbol{X}_m(j)$ is the maximum absolute element-wise difference of them. With a preset threshold r, B_i is the number of $\boldsymbol{X}_m(j)$ such that $d(\boldsymbol{X}_m(i), \boldsymbol{X}_m(j)) \le r$. Let $\Phi^m(r) = (n-m+1)^{-1} \sum_{i=1}^{n-m+1} \log(B_i / (n-m+1))$. The ApEn is defined as $ApEn = \Phi^m(i) - \Phi^{m+1}(i)$. Larger ApEn means that the time series is more irregular or smaller chance of repeated template sequences in the time series. However, ApEn overestimates the similarity since it counts $\boldsymbol{X}_m(i)$ itself when finding B_i, which is called the self-match [richman00]. The SampEn was designed to reduce the bias caused by self-matching. The SampEn of HRV has been shown to be useful to detect sepsis in neonatial intensive care unit [Lake02].

The interactions between physiological time series can also be sketched by the permutation entropy, which is the complexity of a symbolic sequence after embedding the time series into a high dimensional space. First we need to define the ordinal pattern. Given a sequence of numeric elements $x_1, x_2, \ldots x_n$, its ordinal pattern is the permutation $\pi = (i_1, i_2, \ldots i_n)$ that sorts the elements in ascending order such that $x_{i_1} \le x_{i_2} \le \ldots \le x_{i_n}$. Given a time series $x_{1 \ldots N}$ and a time window of length L, the order L permutation entropy can be 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),

$$H_L = -\sum_{\pi_k \in S_L} P(\pi_k) \log\big(P(\pi_k)\big), \quad (\#.4)$$

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

$$R_L^{\alpha} = \frac{1}{1-\alpha} \log\big(\sum_{\pi_k \in S_L} P(\pi_k)^{\alpha}\big). \quad (\#.5)$$

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.

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.

#.4 Machine learning framework and its applications

In the medical field, statistical models are traditionally used to test casual hypotheses, with the purpose focused on explanation [Shmueli10]. Consider a new surgical procedure is being tested for its effect of improving patients' long term neurologic outcome. Doctors may firstly form a statistical null hypothesis that the new procedure has no statistical significant improvement of patients' long term neurologic outcome compared with the old procedure.

Sample size and confounding factors will be identified, before the randomized case enrollment. With all collected data, appropriate statistical testing method will be used to test if the null hypothesis should be rejected at a certain significance level. There are many other statistical models to explain the association between factors and outcomes, such as multivariate regression and Bayesian models. No matter what are the approaches, model interpretability is always emphasized in the explanatory studies.

Another type of purpose is to predict. In daily patient care, clinicians also want to know, for example, if a patient of blunt injury needs massive transfusion after admission, or if a TBI patient will need decompressive craniectomy procedure after 6 hours. With all available data already observed for those patients, reliable prediction can assist clinicians to make early decisions, such as calling the blood bank to prepare sufficient amount of blood product, or early schedule the use of the operating room and notify the operation team. As the decisions based on the predictions have immediate consequences and cost, the prediction must be of high accuracy. In such situation, interpretability could be sacrificed for high prediction accuracy.

Machine learning methods are a collection of algorithms that can discover patterns from data and use them to predict future values. They can be roughly categorized as supervised, unsupervised, and semi-supervised learning. Usually, there are pre-processing steps to convert raw data into a feature set that characterize the observed object or task. In the past, an experienced clinician distills concise and practical rules from years of observation and clinical practice. With massive data, such a process can be accelerated with automated machine learning algorithms. However, the algorithms may generate counterintuitive models, misled by outliers, missing values, biased data, and incorrect assumptions [Bishop06; Lantz13]. Therefore, *a priori* knowledge and model validation are essential in building and selecting models. Moreover, a machine learning method could learn from a training dataset

and attempt to minimize its error on that set. For practical use, we hope the learned model also makes small errors on previously unseen data. A prediction model that can perform well on new data has good generalization ability.

For comprehensive learning algorithms, interested readers can find useful information from Murphy, Bishop, or Hastie's books and their references [murphy12, Bishop06, Hastie11]. In this section, we discuss more on methods that increase the model generalization and the metrics of evaluating the models. With those tools, we could better validate the models and create actionable models.

#.4.1 Feature selection

Instead of using raw data as inputs, machine learning methods often use variables that characterize the data or learning task's properties (features). Feature selection is an important part of data pre-processing before building a simple and robust model. In many applications, including the Big Data scenario, we can find many attributes to characterize the data, without knowing which ones may contribute to the improved accuracy of prediction. Features of little or no importance not only increase the model complexity but also have negative effect on parameter estimation for those features that are of importance to the prediction [Kuhn13]. With massive data, such overfitting is more likely to happen than underfitting.

One class of feature selection methods are model-independent, which are called the filter type methods. They evaluate features according to a certain criterion, before adopting any learning algorithm. There are various criteria that can be used. For example, zero or near zero variance features are often excluded. This is convenient and works in most situations. However, we should be careful, if there is evidence showing such near zero variance features may have strong prediction of the outcome. For example, a certain type of drug, like barbiturate, is only used as a last resort when patients are in critical condition. If one binary feature indicates if a patient was administrated barbiturate in hospital, it may have almost

near zero variance since most patients would not receive this drug. But this feature could be a strong predictor of unfavorable outcome, such as mortality.

Some filter type feature selection methods use the known outcome in the training set. For classification problems, one feature selection approach often seen from medical literature is the use of Receiver Operating Characteristic (ROC) analysis. Besides ROC, information gain can also be used to evaluate each feature's importance to decrease the uncertainty about the outcome. For continuous value prediction, correlation could be used to find a subset of features that each has strong correlation with the outcome while this subset of features are weakly correlated internally. Those methods have low computational complexity, which grows linearly with the number of features. Since the selection is independent of the learning algorithm, the filter type feature selection methods are good for initial fast screening of large amount of features to filter out possibly non-important ones. Also, the ignorance of interaction between features may result in redundant selection.

Another class of feature selection methods search the important features within the context of models, which are call the wrapper type methods. Given a learning method and a performance metric, the wrapper methods search an optimal subset of the features that maximizes model performance. Arguably, stepwise selection is the most frequently used wrapper type feature selection method. In the forward selection step, each feature that is not in the current model is added to a temporary model. The new model is evaluated based on a certain criteria. For example, the most debated criterion is the use of p-value as the inclusion condition. If the hypothesis test calculates the p-value of the newly added term is less than the inclusion threshold, then keep the feature in the next around evaluation. Other inclusion criteria include the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC), which are considered better than the p-value based selection. The search halts when there are no features that meet the inclusion criteria remain outside the model. In the

backward selection step, the full model that contains all candidate features, or all selected features from the forward selection, has features iteratively removed based on the exclusion criteria, which could be p-value, AIC or BIC, etc..

Permutation is another way to evaluate feature importance in a model. Intuitively, if a feature is of real importance, then randomly permuting its values will greatly reduce the model performance. Given a trained model and performance metric, the trained baseline model has its performance P_b. Each feature is randomly permuted while other features remain unchanged. The model is then trained again with the altered ith feature and evaluated for a new performance P_{ri}. The change of performance metric is calculated as $\Delta P_i = |P_b - P_{ri}|$. To have a robust estimation of the performance change, we can repeate such random permutation many times, and use the averaged difference. In this way, we can rank all features according to the change of performance metric. Then the selected features are optimal in term of that metric.

The third class of feature selection methods train the model and select features simultaneously, which are called the embedded methods. Many machine learning algorithms have such inherent feature selection mechanism, such as random forest, relevance vector machine, decision trees. A more general approach is to add a penalty term for the model complexity to the cost function using Lagrangian multiplier. Given a dataset $D = \{(\boldsymbol{x}_i, y_i)\}$, where $\boldsymbol{x} \in R^d$ and $y \in R$. A learning method finds the optimal parameter set $\boldsymbol{w}$ through minimizing a loss function f by $\min_{w} f(\boldsymbol{x}, y, \boldsymbol{w})$. Through a Lagrangian multiplier the optimization target can be extended to be

$$\min_{w} f(\boldsymbol{x}, y, \boldsymbol{w}) + \lambda \|\boldsymbol{w}\|_1, \qquad (\#.6)$$

where $\lambda > 0$ is called the regularization parameter. The penalty term $\|\boldsymbol{w}\|_1$ is the ℓ_1 norm of the model parameter set $\boldsymbol{w}$. The equation (#.6) is well known as the Least Absolute

Shrinkage and Selection Operator (LASSO) [Tibshirani96]. It has a nice property to suppress variable coefficients, which introduces sparsity while learning the parameter set. It has been shown to be efficient to create parsimonious and robust models [Hastie15, Murphy12].

All the three classes of feature selection methods have their own applicable scenarios. For the filter type, they are computationally efficient. Because they can be independent of the model, during the initial stage of experiment design in medical studies, those methods can be used to quickly filter out possibly useless variables, even if there are no known outcome labels yet. This can help reduce the amount of expensive data collection. However, filter methods lack knowledge of the interaction between features, thus may include redundant features. The wrapper type methods iteratively search the optimal subset of features by testing each variable 'contribution' to a selected model performance metric. The search scheme takes variables' interactions into consideration. But each evaluation iteration requires the model to be trained again, which could be time consuming. The embedded methods have much more desired advantages. First, they include feature interaction for consideration. Second, with efficient optimization solvers, the parameter learning and regularization can be done at the same time efficiently. Moreover, through special design of the regularization terms, not only feature importance but also the structure among feature groups could be selected [Yuan06].

#.4.2 Performance metrics

An appropriate model performance metric allows defining right learning objectives, and can increase the chance of building generative models. Two levels of performance are crucial in creating a good predictive model. First, we need to evaluate the model's generalization capability, namely, can we have a high expectation on its performance on future unseen data? Second, we want to know the model's performance on the training data, namely, how well this model has learned from what it could observe?

In practice, we always have finite observations. One reason is that collecting data could be expensive. Another reason is that it is infeasible to sample the entire population in most situations. To know models' expected performance on unseen data, we have mainly two approaches called external or internal validations. The former uses data from external sources, such as from other geologically different clinical facilities, or from other historical time points. The latter reserves an internal small portion of the collected data and sets it aside only for testing.

External validation is often desired in clinical studies. Data collected from one regional hospital may still have its sampling bias, which are mainly influenced by its local demographics. Also, data collected from a civilian hospital may not represent the military population. Hence, it is important to validate models built from single-center studies for their generalization. There are many large-scale, high-quality clinical databases maintained for public use. PhysioNet is a large physiological database archiving data contributed from worldwide [Moody09]. The National Trauma Data Bank (NTDB) is the largest U.S. trauma registry dataset, which has been used in study trauma injury epidemiology, validation of guidelines from clinical organizations [Webman16,Tinkoff08,Millham04].

Despite being in the Big Data era, collecting large amounts of clinical data is still expensive and requires tremendous amount effort to store and process massive data. Hence, it is luxury to keep a large part of data outside of training. The internal validation is necessary for most single-center studies, especially when external data are not available. The internal validation randomly partitions the collected data to be training and testing. A typical validation scheme is the k-fold cross validation. Given a dataset, it is randomly partitioned into roughly k equal size non-overlapping subsets. In the ith validation, where $i = 1 \ldots k$, the ith subset is used as testing set and the remaining data are used for training. Such process is iterated over all k subsets. After one round of k-fold cross validation, there are k model

evaluations. Often, we can repeat such k-fold cross validation N times by randomly partition the set again. With the $k \times N$ evaluations, we can use their averaged performance or other robust statistic (e.g. standard deviation, median, etc.) to estimate the model's future performance on unseen data. Such random subset sampling is a simulation of possibly different distribution of new data.

Commonly, mean squared error (MSE) or root mean square error (RMSE) is used in evaluating continuous outcomes. Accuracy or the confusion matrix is typically used for discrete outcomes. When the prediction problems have some special issues, such as imbalance in data labels, or different preference on incorrectly predicted cases, we need to carefully select the performance metrics. In this section, we discuss the receiver operating characteristic (ROC) curve and the precision and recall curve (PRC), the two types of model performance metrics that are widely used in medical classification problems.

Accuracy as a model performance metric has some drawbacks in medical data analysis. It could be misleading when the data labels (e.g. positive and negative outcomes in binary classification) are highly imbalanced. A classifier could cheat to achieve high accuracy by just predicting all instances to be the most frequent class label. Unfortunately, in many medical problems, the outcome of interest often has small portion. For example, in a regional level 1 trauma center massive transfusion rate could be as low as 1.3%~2.2% in adult patients [Nehu16]. A classifier could predict no massive transfusion for all patients and achieve >98% accuracy. However, miss predicting for a patient who needs massive transfusion may result in severe clinical and social outcome.

To alleviate this problem, accuracy is broken into four parts for binary classifications, which is often described as the confusion matrix. The correctly predicted positive cases are true positives (TP). The incorrectly predicted positive cases are false positives (FP). Similarly, the true negatives (TN) are correctly predicted negative cases; and the false

negative (FN) are incorrectly predicted negative cases [Fawcett03, Krzanowski09].

Given a set of instances for prediction, a classifier gives corresponding predicted values, or called prediction scores. If we sort all the scores in ascending order and use each unique value as a threshold, all cases with score values smaller than the threshold are classified as negative, whereas those are higher will be classified as positive. Then for each threshold, we can calculate the false positive rate (FPR) and the true positive rate (TPR), which constitutes a point in a 2D coordinates. After iterating all possible thresholds, a full curve is generated, called the ROC curve. The PRC curve is created in a similar way, with positive predicted values (PPV=TP/(TP+FP), also known as precision), and true positive rate (TPR, also known as recall).

The ROC and PRC curves provide a full spectrum evaluation of a classifier by visualizing FPR, TPR, or PPV, TPR for all possible thresholds, instead of giving a single decision point. The ROC curve is widely used to evaluate prediction models in many medical studies, while the PRC curve is much less frequently seen. Davis et al. [Davis06] proved that ROC and PRC curves have one-to-one correspondence, given a dataset with finite positive and negative cases. If one ROC curve has all its points above or equal to another ROC curve, it is said that the first ROC curve dominates the second one; and this is true if and only if the first PRC curve dominates the second PRC curve. However, in many situations, ROC curves from multiple models built from the same dataset are twined or close to each other. When the dataset is highly imbalanced, i.e. the negative cases significantly outnumber the positive cases, the PRC curve has better ability to identify classifiers that have good performance in predicting the positive cases.

From one of our projects, we collected 1191 injured patients, and collected continuous physiological vital signs to predict the next 3 hour blood product use. Among the 1191 patients, only 7.2% received blood product. As an illustration, Figure 7 compares the ROC

and PRC curves of three methods we used. To focus on the use of the performance curves, we skip the details of the models here. As we can see from Figure 7a, the blue ROC curve (Method 1) dominates the other two, and so does it in the PRC curves. The red and green ROC curves (Methods 2 and 3) are close to each other and intertwined. From Figure 7a, the two methods have very similar performance, but from Figure 7b, we can observe that the green PRC curve (Method 3) has higher precision than the red PRC curve (Method 2). Therefore, in comparing the methods, we may consider the Methods 1 and 3 have better performance than the Method 2.

#.5 Computing issues

Storing, processing, and learning from massive medical data require intensive calculation. First, having a high performance file format to store and organize large amounts of data is critical for the input and output (I/O) of data processing. For a typical TBI patient staying 7 days in a trauma centre, five 240-Hz waveforms are monitored and up to 700 million data points (equivalent to 1-gigabyte disk size, if data are stored in 12-bit format) would be collected. Traditional spreadsheet based data management becomes less efficient within such Big Data scenario. Hierarchical Data Format (HDF) is a high performance data format that offers on-the-fly data compression and high I/O performance. It also supports to read from or write to a subset of a dataset, without loading the entire data file into memory.

Second, utilizing the independence among tasks allows for parallel data processing, thus making full use of multi-core or many-core machines. There are mainly two levels of parallelism in typical medical prediction model training. One is between subjects and another is within subject. Often, the feature extraction from a patient's data is independent of others. At this level, we can distribute study cases evenly to all computing units. Within each subject, many tasks can also be done simultaneously. For example, features derived from single variables can be calculated on separate cores. Features from moving windows are also highly

parallelizable. In the model learning steps, repeated cross-validation is commonly adopted, to test and validate these models' performance on new data and to prevent potential over-fitting. A balanced training and testing model prediction is used to see if the model can be generalized to new previously unused data. For example, with multiple combinations of five outcomes, six feature groups, 10-fold cross-validation repeated ten times, about 1500 to 3000 multiples of model calculations and 100 to 300 model comparisons and statistical tests are required. Parallel training and testing can be used to speed up the learning process.

Many programming languages and scientific data analysis libraries support parallel computing, such as OpenMP (Open Multi-Processing), MPI (Message Passing Interface) for CPU parallelism, and CUDA for graphics processing unit (GPU) parallelism. Moreover, nowadays, cloud computing services provide scalable on-demand use of shared computing resources, including CPUs/GPUs, memory, storage and security. The end users from hospital or medical research institutes can be freed from the burden of building and maintaining expensive high performance data processing equipments and infrastructures. Giant vendors, such as AWS (Amazon Web Services), Microsoft Azure, and Google Cloud Platform (GCP), are finding more and more applications in medical data mining and machine learning.

#.6 Discussion

Unprecedented volume of data is generated daily in trauma patient care. However, one cruel fact is that the health caring resources are still very limited in both field and hospital. Matched blood product, operation rooms, experienced healthcare providers such as surgeons, anesthesiologist, and nurses are always scarce. The ultimate goal of Big Data in trauma patient care is to intelligently optimize the allocation of limited healthcare resources, by reliable prediction for needs of life-saving interventions, and early decision of therapeutic plans. With automated data processing and informative data aggregation, useful knowledge from massive data can be used by clinicians in a simple way for decision making or

prioritizing care in the busy hospital environment.

In this chapter, we reviewed a few critical components in large-scale medical data analysis, including reliable data collection, feature extraction, feature selection, and model evaluation. To distill reliable predictive models from massive medical data, generalizable good performance on unseen data and interpretability are the top two most important factors to consider. Clinical expert knowledge could assist to design meaningful and useful features that may associate with outcomes of special interest, such as transfusion, use of operation room, length of stay, and other actionable life saving interventions. Moreover, thorough testing with internal and external data provides some evidence of how the models may perform, before the learned models could be deployed.

Figure Legend

Figure 1. The MoMs system architecture with triple modular redundancy design using three BedMaster servers.

Figure 2. A portion of MoMs viewer for data collection status. Green cells (shown): collection is active (within last 5 minutes); yellow (not shown): collection was active 5 minutes to 4 hours ago; red (shown): no data collection has occurred in more than 4 hours; gray (shown): a bedside collection is offline. In each cell letter "A" means admitted; letter "D" discharged. The pink background cell indicates a patient with an intracranial monitor in place; ICP value appears in white.

Figure 3. An exemplary ECG segment with identified P,Q,R,S,T peaks. Five items from the segment are used for ECG feature calculation. Item 1 is the NN interval. Items 2 and 4 are Q to R rising time and R to S falling time. Items 3 and 5 are Q to R rising amplitude and R to S falling amplitude.

Figure 4. Illustration of ECG, PPG segments of bad and good signal quality, evaluated by z-test on NN intervals. On left side, red horizontal bars flag ECG, PPG segments of bad signal quality.

Figure 5. (Top) An exemplary PPG segment with identified peaks and valleys. (Middle) the first derivative of PPG signal. (Bottom) the second derivative of PPG signal. Item 1 is the peak-peak interval. Items 2 is the valley-peak rising time. Item 3 is the valley-peak rising amplitude. Items 4 and 5 are peak-valley falling time and amplitude. Item 6 is the notch area, where the time duration and amplitude change can be calculated.

Figure 6. Fourier transform of 10 minutes ICP waveform. Pulse wave components shows four harmonics, with the first one corresponding to heart rate. The other two components, respiratory wave and slow waves are both identifiable from the frequency domain.

Figure 7a. Example ROC curves of three methods.

Figure 7b. Example PRC curves of three methods.

References:

[Aaslid89]. Aaslid, Rune, Kari-Fredrik Lindegaard, Wilhelm Sorteberg, and Helge Nornes. "Cerebral autoregulation dynamics in humans." *Stroke* 20, no. 1 (1989): 45-52.

[Acharya06]. Acharya, U. Rajendra, K. Paul Joseph, Natarajan Kannathal, Choo Min Lim, and Jasjit S. Suri. "Heart rate variability: a review." *Medical and biological engineering and computing* 44, no. 12 (2006): 1031-1051.

[Allen07]. Allen, John. "Photoplethysmography and its application in clinical physiological measurement." *Physiological measurement* 28, no. 3 (2007): R1.

[Aslanidia15]. Aslanidia T. Perspectives of Autonomic Nervous System Perioperative Monitoring- focus on selected tools. International Archives of Medicine. 2015; 8(22): 1-7.

[Baguley06]. Baguley, Ian J., Roxana E. Heriseanu, Kim L. Felmingham, and Ian D. Cameron. "Dysautonomia and heart rate variability following severe traumatic brain injury." *Brain Injury* 20, no. 4 (2006): 437-444.

[Balestreri04]. Balestreri, M., M. Czosnyka, L. A. Steiner, E. Schmidt, P. Smielewski, B. Matta, and J. D. Pickard. "Intracranial hypertension: what additional information can be derived from ICP waveform after head injury?." *Acta neurochirurgica* 146, no. 2 (2004): 131-141.

[Bengio15]. Bengio, Yoshua, Ian J. Goodfellow, and Aaron Courville. "Deep learning." An MIT Press book in preparation. Draft chapters available at http://www. iro. umontreal. ca/~bengioy/dlbook (2015).

[Bishop06]. Bishop CM (2006) *Pattern Recognition and Machine Learning*. Springer, New York.

[Brown14]. Brown, J. B., Gestring, M. L., Forsythe, R. M., Stassen, N. A., Billiar, T. R., Peitzman, A. B., & Sperry, J. L. (2015). Systolic blood pressure criteria in the National Trauma Triage Protocol for geriatric trauma: 110 is the new 90. *The journal of trauma and*

acute care surgery, *78*(2), 352.

[BTI guide07]. Brain Trauma Foundation, American Association of Neurological Surgeons; Congress of Neurological Surgeons. Guidelines for the management of severe head injury. J Neurotrauma. 2007;24(Suppl 1):S1-S106.

[Cover12]. Cover, Thomas M., and Joy A. Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[Czosnyka96]. Czosnyka, Marek, Piotr Smielewski, Peter Kirkpatrick, David K. Menon, and John D. Pickard. "Monitoring of cerebral autoregulation in head-injured patients." *Stroke* 27, no. 10 (1996): 1829-1834.

[Czosnyka14]. Czosnyka, Marek, and John D. Pickard. "Monitoring and interpretation of intracranial pressure." *Journal of Neurology, Neurosurgery & Psychiatry* 75, no. 6 (2004): 813-821.

[Davis06]. Davis, Jesse, and Mark Goadrich. "The relationship between Precision-Recall and ROC curves." In *Proceedings of the 23rd international conference on Machine learning*, pp. 233-240. ACM, 2006.

[DuBose11]. DuBose, Joseph J., Gallinos Barmparas, Kenji Inaba, Deborah M. Stein, Tom Scalea, Leopoldo C. Cancio, John Cole, Brian Eastridge, and Lorne Blackbourne. "Isolated severe traumatic brain injuries sustained during combat operations: demographics, mortality outcomes, and lessons to be learned from contrasts to civilian counterparts." *Journal of Trauma and Acute Care Surgery* 70, no. 1 (2011): 11-18.

[Drongelen06]. Van Drongelen, Wim. *Signal processing for neuroscientists: an introduction to the analysis of physiological signals*. Academic press, 2006.

[Elgendi14]. Elgendi, Mohamed, Ian Norton, Matt Brearley, Derek Abbott, and Dale Schuurmans. "Detection of a and b waves in the acceleration photoplethysmogram." *Biomedical engineering online* 13, no. 1 (2014): 1.

[Ernst14]. Ernst, Gernot. *Heart rate variability*. London: Springer, 2014.

[Excel13]. Excel Medical Electronics, LLC, BedMasterEx Operator's Manual, Jupiter, FL, 2013.

[Fawcett03]. Fawcett, Tom. "ROC graphs: Notes and practical considerations for researchers." *Machine learning* 31, no. 1 (2004): 1-38.

[Garg05]. Garg, Amit X., Neill KJ Adhikari, Heather McDonald, M. Patricia Rosas-Arellano, P. J. Devereaux, Joseph Beyene, Justina Sam, and R. Brian Haynes. "Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review." *Jama* 293, no. 10 (2005): 1223-1238.

[Goldstein96]. Goldstein, Brahm, Mark H. Kempski, Donna BA DeKing, Christopher Cox, David J. DeLong, Mary M. Kelly, and Paul D. Woolf. "Autonomic control of heart rate after brain injury in children." *Critical care medicine* 24, no. 2 (1996): 234-240.

[Goldberger00]. Goldberger, Ary L., Luis AN Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. "Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiological signals." *Circulation* 101, no. 23 (2000): e215-e220.

[Goodman13]. Goodman, Brent, Bert Vargas, and David Dodick. "Autonomic Nervous System Dysfunction in Concussion (P01. 265)." *Neurology* 80, no. 7 Supplement (2013): P01-265.

[Hastie11]. Trevor J.. Hastie, Robert John Tibshirani, and Jerome H. Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2011.

[Hastie15]. Hastie, Trevor, Robert Tibshirani, and Martin Wainwright. *Statistical learning with sparsity: the lasso and generalizations*. CRC Press, 2015.

[Holcomb10]. Holcomb, John B. "Optimal use of blood products in severely injured trauma

patients." *ASH Education Program Book* 2010, no. 1 (2010): 465-469.

[Kahraman10]. Kahraman, S., Dutton, R. P., Hu, P., Xiao, Y., Aarabi, B., Stein, D. M., & Scalea, T. M. (2010). Automated measurement of "pressure times time dose" of intracranial hypertension best predicts outcome after severe traumatic brain injury. *Journal of Trauma and Acute Care Surgery*, *69*(1), 110-118.

[Kahraman10b]. Kahraman, Sibel, Richard P. Dutton, Peter Hu, Lynn Stansbury, Yan Xiao, Deborah M. Stein, and Thomas M. Scalea. "Heart rate and pulse pressure variability are associated with intractable intracranial hypertension after severe traumatic brain injury." *Journal of neurosurgical anesthesiology* 22, no. 4 (2010): 296-302.

[Kamath12].Kamath, Markad V., Mari Watanabe, and Adrian Upton, eds. *Heart rate variability (HRV) signal analysis: clinical applications*. CRC Press, 2012.

[Kashif12]. Kashif, Faisal M., George C. Verghese, Vera Novak, Marek Czosnyka, and Thomas Heldt. "Model-based noninvasive estimation of intracranial pressure from cerebral blood flow velocity and arterial pressure." *Science translational medicine* 4, no. 129 (2012): 129ra44-129ra44.

[Khawaja06]. Khawaja, Antoun. *Automatic ECG analysis using principal component analysis and wavelet transformation*. Univ.-Verlag Karlsruhe, 2006.

[Kipnis12]. Kipnis, Eric, Davinder Ramsingh, Maneesh Bhargava, Erhan Dincer, Maxime Cannesson, Alain Broccard, Benoit Vallet, Karim Bendjelid, and Ronan Thibault. "Monitoring in the intensive care." *Critical care research and practice* 2012 (2012).

[Koht12]. Koht, Antoun, Tod B. Sloan, and J. Richard Toleikis. *Monitoring the nervous system for anesthesiologists and other health care professionals*. New York, NY, USA:: Springer, 2012.

[Krzanowski09]. Krzanowski, Wojtek J., and David J. Hand. *ROC curves for continuous data*. CRC Press, 2009.

[Kuhn13]. Kuhn, Max, and Kjell Johnson. *Applied predictive modeling*. New York: Springer, 2013.

[Lake02]. Lake, Douglas E., Joshua S. Richman, M. Pamela Griffin, and J. Randall Moorman. "Sample entropy analysis of neonatal heart rate variability." *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 283, no. 3 (2002): R789-R797.

[Lantz13]. Lantz, Brett. *Machine learning with R*. Packt Publishing Ltd, 2013. Birmingham.

[Lehman14]. Li-wei, H. Lehman, Ryan P. Adams, Louis Mayaud, George B. Moody, Atul Malhotra, Roger G. Mark, and Shamim Nemati. "A physiological time series dynamics-based approach to patient monitoring and outcome prediction." *IEEE journal of biomedical and health informatics* 19, no. 3 (2015): 1068-1076.

[Liu14]. Liu, N. T., Holcomb, J. B., Wade, C. E., Darrah, M. I., & Salinas, J. (2014). Utility of vital signs, heart rate variability and complexity, and machine learning for identifying the need for lifesaving interventions in trauma patients. *Shock*, *42*(2), 108-114.

[Lu08]. Lu, Sheng, He Zhao, Kihwan Ju, Kunson Shin, Myoungho Lee, Kirk Shelley, and Ki H. Chon. "Can photoplethysmography variability serve as an alternative approach to obtain heart rate variability information?." *Journal of clinical monitoring and computing* 22, no. 1 (2008): 23-29.

[Mackenzie14]. Mackenzie, C. F., Wang, Y., Hu, P. F., Chen, S. Y., Chen, H. H., Hagegeorge, G., ... & ONPOINT Study Group. (2014). Automated prediction of early blood transfusion and mortality in trauma patients. *Journal of Trauma and Acute Care Surgery*, *76*(6), 1379-1385.

[Milham04]. Millham, Frederick H., and Wayne W. LaMorte. "Factors associated with mortality in trauma: re-evaluation of the TRISS method using the National Trauma Data

Bank." *Journal of Trauma and Acute Care Surgery* 56, no. 5 (2004): 1090-1096.

[Moody09]. Moody, George B. "PhysioNet: Research Resource for Complex Physiological Signals." http://ecg.mit.edu/george/publications/physionet-jecg-2009.pdf, Retrieved at Jul. 15, 2016.

[Murphy12]. Murphy, Kevin P. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[Nehu16]. Parimi, Nehu, Peter F. Hu, Colin F. Mackenzie, Shiming Yang, Stephen T. Bartlett, Thomas M. Scalea, and Deborah M. Stein. "Automated continuous vital signs predict use of uncrossed matched blood and massive transfusion following trauma." *Journal of Trauma and Acute Care Surgery* 80, no. 6 (2016): 897-906.

[Nitzan97]. Nitzan, M., Babchenko, A., Khanokh, B., & Landau, D. (1998). The variability of the photoplethysmographic signal-a potential method for the evaluation of the autonomic nervous system. *Physiological measurement*, *19*(1), 93.

[Olaussen14]. Olaussen A, Blackburn T, Mitra B, Fitzgerald M. Review article: Shock Index for prediction of critical bleeding post-trauma: A systematic review. *Emergency Medicine Australasia*. June 2014;26(3):223-228.

[Palmer10]. Palmer, Ronald W. *Integrated diagnostic and treatment devices for enroute critical care of patients within theater*. No. RTO-MP-HFM-182. ARMY MEDICAL RESEARCH AND MATERIEL COMMAND FORT DETRICK MD, 2010.

[Pan85]. Pan, Jiapu, and Willis J. Tompkins. "A real-time QRS detection algorithm." *IEEE transactions on biomedical engineering* 3 (1985): 230-236.

[Perkin12]. Perkins JG, Beekley AC (2012) Damage control resuscitation. In: Savitsky E, Eastridge B, eds. *Combat Casualty Care: Lessons Learned from OEF and OIF*. Department of the Army, Office of the Surgeon General, Borden Institute, Washington, DC: 121-64.

[Picus91]. Pincus, Steven M. "Approximate entropy as a measure of system complexity."

Proceedings of the National Academy of Sciences 88, no. 6 (1991): 2297-2301.

[Press07]. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., Chapter 14: Statistical Description of Data. In: Numerical Recipes the Art of Scientific Computing, 3rd, Cambridge Univeristy Press, 2007.

[Provost13]. Provost, Foster, and Tom Fawcett. "Data science and its relationship to big data and data-driven decision making." *Big Data* 1, no. 1 (2013): 51-59.

[Reddy15]. Reddy, Chandan K., and Charu C. Aggarwal, eds. *Healthcare data analytics*. Vol. 36. CRC Press, 2015.

[Richman00]. Richman, J. S., & Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, *278*(6), H2039-H2049.

[Saeed11]. Saeed, Mohammed, Mauricio Villarroel, Andrew T. Reisner, Gari Clifford, Li-Wei Lehman, George Moody, Thomas Heldt, Tin H. Kyaw, Benjamin Moody, and Roger G. Mark. "Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): a public-access intensive care unit database." *Critical care medicine* 39, no. 5 (2011): 952.

[Sasser11] Sasser, S. M., Hunt, R. C., Faul, M., Sugerman, D., Pearson, W. S., Dulski, T., ... & Lerner, E. B. (2012). Guidelines for field triage of injured patients: recommendations of the National Expert Panel on Field Triage, 2011. *MMWR. Recommendations and reports: Morbidity and mortality weekly report. Recommendations and reports/Centers for Disease Control*, *61*(RR-1), 1-20.

[Schmidt15]. Schmidt, Paul E., Paul Meredith, David R. Prytherch, Duncan Watson, Valerie Watson, Roger M. Killen, Peter Greengross, Mohammed A. Mohammed, and Gary B. Smith. "Impact of introducing an electronic physiological surveillance system on hospital mortality." *BMJ quality & safety* (2014): bmjqs-2014.

[Shmueli10]. Shmueli, Galit. "To explain or to predict?." *Statistical science* (2010): 289-310.

[Shooman02]. Shooman, Martin L. *Reliability of computer systems and networks: fault tolerance, analysis, and design*. John Wiley & Sons, 2003.

[Stein11]. Stein, Deborah M., Peter F. Hu, Megan Brenner, Kevin N. Sheth, Keng-Hao Liu, Wei Xiong, Bizhan Aarabi, and Thomas M. Scalea. "Brief episodes of intracranial hypertension and cerebral hypoperfusion are associated with poor functional outcome after severe traumatic brain injury." *Journal of Trauma and Acute Care Surgery* 71, no. 2 (2011): 364-374.

[Stein12]. Stein, Deborah M., Peter F. Hu, Hegang H. Chen, Shiming Yang, Lynn G. Stansbury, and Thomas M. Scalea. "Computational gene mapping to analyze continuous automated physiological monitoring data in neuro-trauma intensive care." *Journal of Trauma and Acute Care Surgery* 73, no. 2 (2012): 419-425.

[Stein13]. Stein, Deborah M., Megan Brenner, Peter F. Hu, Shiming Yang, Erin C. Hall, Lynn G. Stansbury, Jay Menaker, and Thomas M. Scalea. "Timing of intracranial hypertension following severe traumatic brain injury." *Neurocritical care* 18, no. 3 (2013): 332-340.

[Steiner02]. Steiner, Luzius A., Marek Czosnyka, Stefan K. Piechnik, Piotr Smielewski, Doris Chatfield, David K. Menon, and John D. Pickard. "Continuous monitoring of cerebrovascular pressure reactivity allows determination of optimal cerebral perfusion pressure in patients with traumatic brain injury." *Critical care medicine* 30, no. 4 (2002): 733-738.

[Takens81]. Takens, Floris. "Detecting strange attractors in turbulence." In *Dynamical systems and turbulence, Warwick 1980*, pp. 366-381. Springer Berlin Heidelberg, 1981.

[TaskForceESC96]. Task Force of the European Society of Cardiology. "Heart rate variability standards of measurement, physiological interpretation, and clinical use." Eur Heart J 17 (1996): 354-381.

[Tibshirani96]. Tibshirani, Robert. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society. Series B (Methodological)* (1996): 267-288.

[Tinkoff08]. Tinkoff, Glen, Thomas J. Esposito, James Reed, Patrick Kilgo, John Fildes, Michael Pasquale, and J. Wayne Meredith. "American Association for the Surgery of Trauma Organ Injury Scale I: spleen, liver, and kidney, validation based on the National Trauma Data Bank." *Journal of the American College of Surgeons* 207, no. 5 (2008): 646-655.

[Voss12]. Voss, A., Goernig, M., Schroeder, R., Truebner, S., Schirdewan, A., & Figulla, H. R. (2012). Blood pressure variability as sign of autonomic imbalance in patients with idiopathic dilated cardiomyopathy. *Pacing and Clinical Electrophysiology*, *35*(4), 471-479.

[Voss15]. Voss, Andreas, Rico Schroeder, Andreas Heitmann, Annette Peters, and Siegfried Perz. "Short-term heart rate variability—influence of gender and age in healthy subjects." *PloS one* 10, no. 3 (2015): e0118308.

[Webman16]. Webman, Rachel B., Elizabeth A. Carter, Sushil Mittal, Jichaun Wang, Chethan Sathya, Avery B. Nathens, Michael L. Nance, David Madigan, and Randall S. Burd. "Association between trauma center type and mortality among injured adolescent patients." *JAMA pediatrics* (2016).

[Yousef12]. Yousef, Q., M. B. I. Reaz, and Mohd Alauddin Mohd Ali. "The analysis of PPG morphology: investigating the effects of aging on arterial compliance." *Measurement Science Review* 12, no. 6 (2012): 266-271.

[Yuan06]. Yuan, Ming, and Yi Lin. Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68, no. 1 (2006): 49-67.

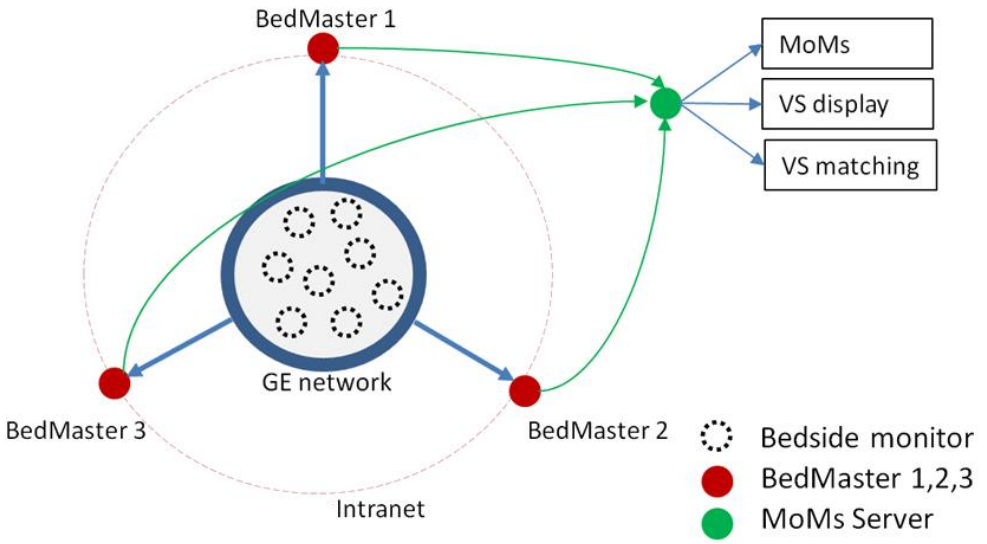

Figure. 1. The MoMs system architecture with triple modular redundancy design using three BedMaster servers.

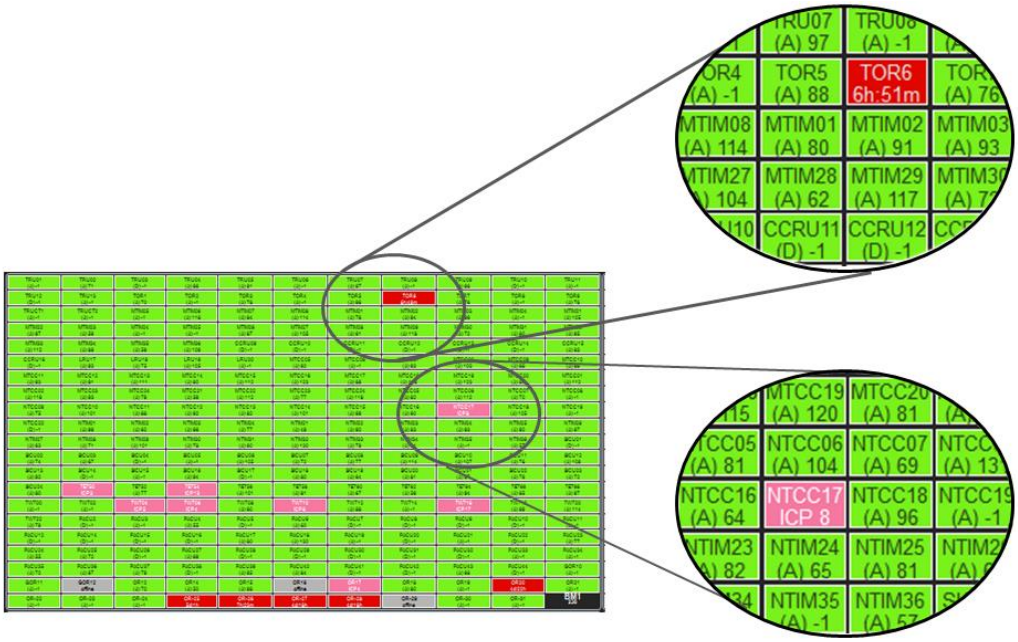

Figure. 2: A portion of MoMs viewer for data collection status. Green cells (shown): collection is active (within last 5 minutes); yellow (not shown): collection was active 5 minutes to 4 hours ago; red (shown): no data collection has occurred in more than 4 hours; gray (shown): a bedside collection is offline. In each cell letter "A" means admitted; letter "D" discharged. The pink background cell indicates a patient with an intracranial monitor in place; ICP value appears in white.

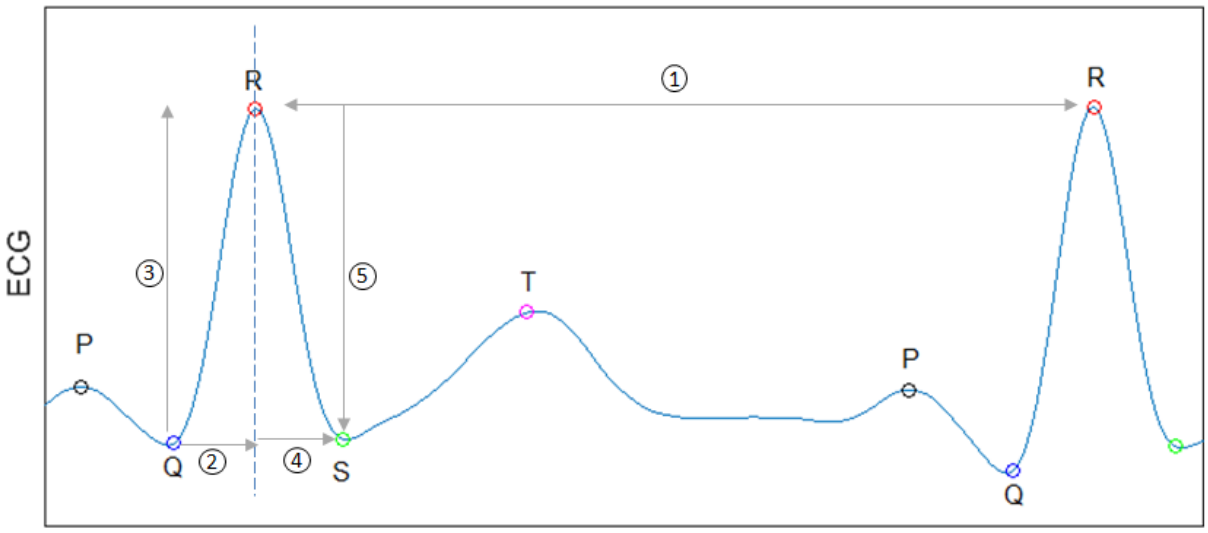

Figure 3. An exemplary ECG segment with identified P,Q,R,S,T peaks. Five items from the segment are used for ECG feature calculation. Item 1 is the NN interval. Items 2 and 4 are Q to R rising time and R to S falling time. Items 3 and 5 are Q to R rising amplitude and R to S falling amplitude.

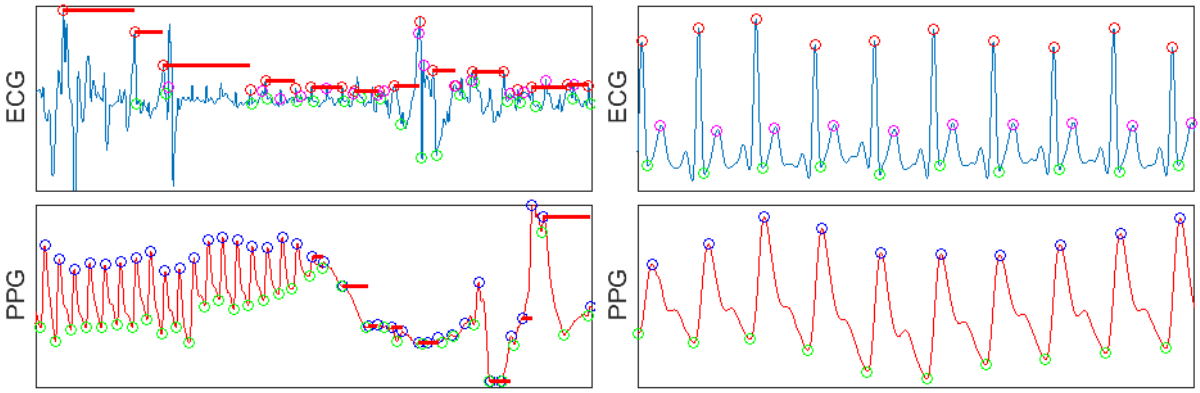

Figure 4. Illustration of ECG, PPG segments of bad and good signal quality, evaluated by z-test on NN intervals. On left side, red horizontal bars flag ECG, PPG segments of bad signal quality.

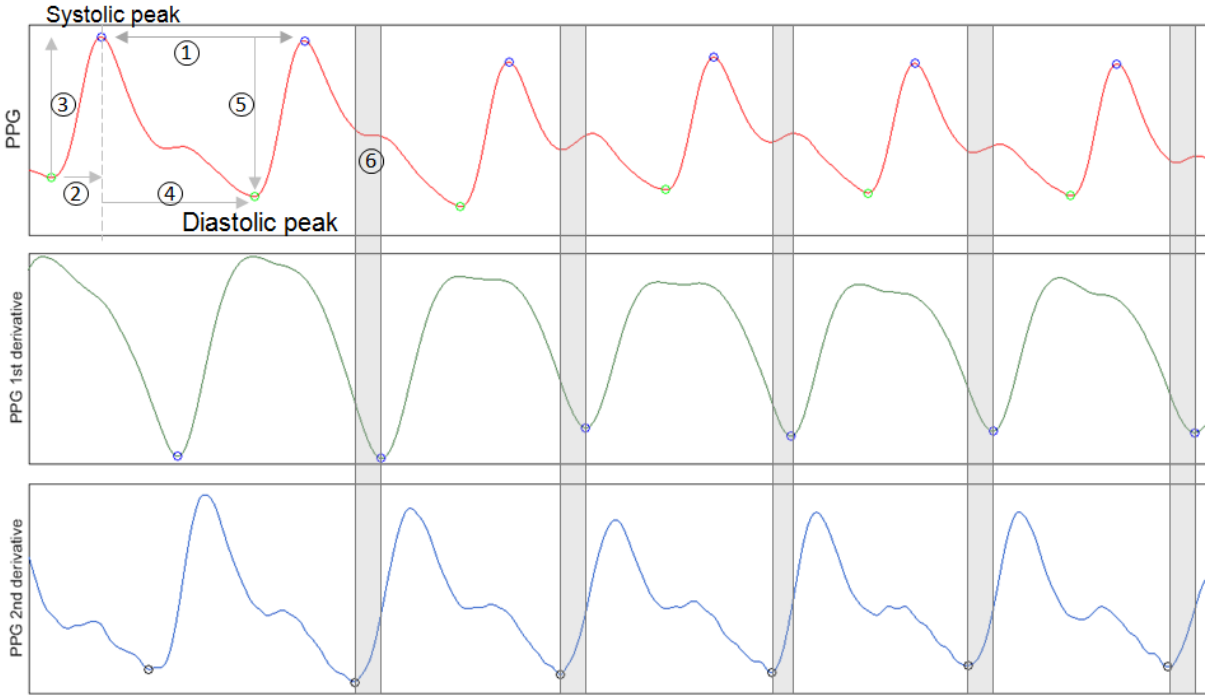

Figure 5. (Top) An exemplary PPG segment with identified peaks and valleys. (Middle) the first derivative of PPG signal. (Bottom) the second derivative of PPG signal. Item 1 is the peak-peak interval. Items 2 is the valley-peak rising time. Item 3 is the valley-peak rising amplitude. Items 4 and 5 are peak-valley falling time and amplitude. Item 6 is the notch area, where the time duration and amplitude change can be calculated.

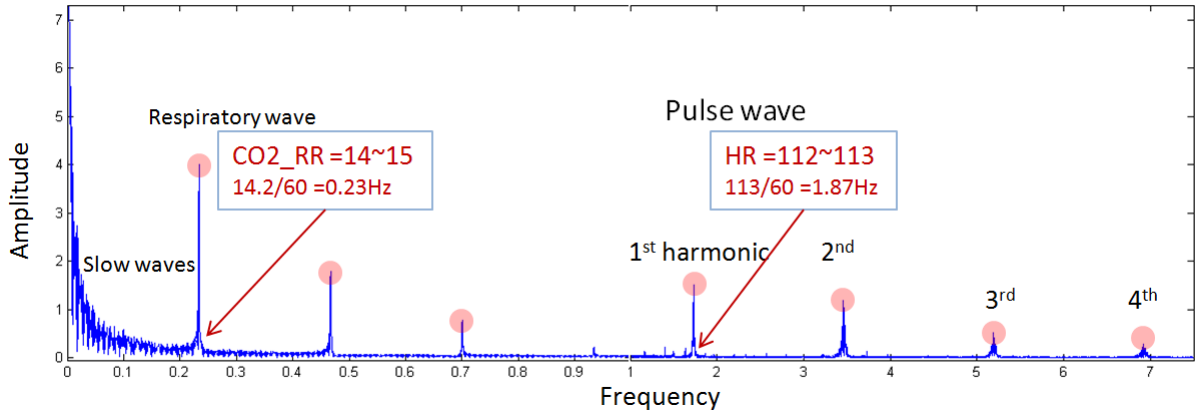

Figure 6. Fourier transform of 10 minutes ICP waveform. Pulse wave components shows four harmonics, with the first one corresponding to heart rate. The other two components, respiratory wave and slow waves are both identifiable from the frequency domain.

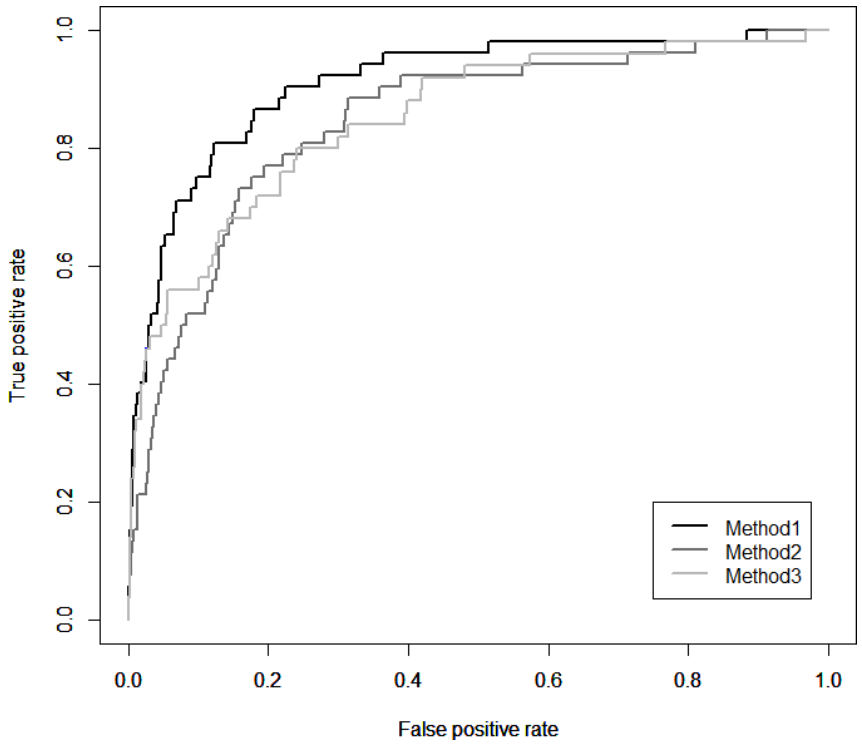

Figure 7a. Example ROC curves of three methods

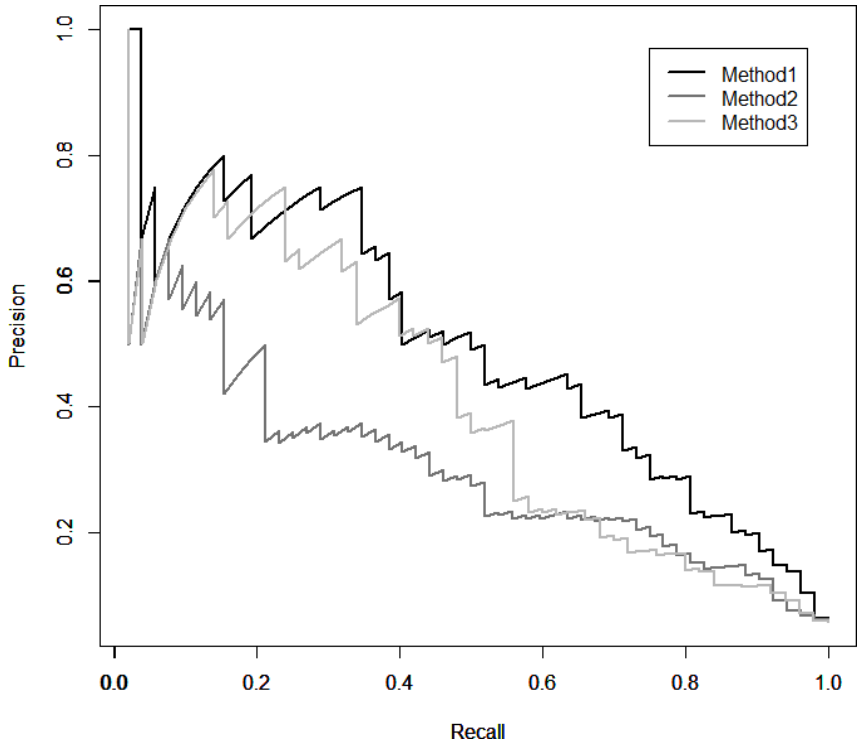

Figure 7b. Example PRC curves of three methods.