Hemodynamic Instability Prediction Through Continuous Multiparameter Monitoring in ICU

Hanqing Cao[1*], Larry J. Eshelman[1], Larry Nielsen[2], Brian D. Gross[2], Mohammed Saeed[2], Joseph J. Frassica[2]
[1]Philips Research North America, Briarcliff Manor, NY 10510 USA
[2]Philips Healthcare, Andover, MA 01810 USA
[*]Corresponding author: Hanqing Cao, Ph. D., Philips Research North America, 345 Scarborough Road, Briarcliff Manor, NY 10510, Phone: 914-945-6153, Fax: 914-945-6580, Email: hanqing.cao@philips.com.

ABSTRACT

Current algorithms identifying hemodynamically unstable intensive care unit patients typically are limited to detecting existing dangerous conditions and suffer from high false alert rates. Our objective was to predict hemodynamic instability at least two hours before patient deterioration while maintaining a low false alert rate, using minute-by-minute heart rate (HR) and blood pressure (BP) data. We identified 66 stable and 104 unstable patients meeting our stability-instability criteria from the MIMIC II database, and developed multi-parameter measures using HR and BP. An instability index combining measures of BP, shock index, rate pressure product, and HR variation was developed from a multivariate regression model to predict hemodynamic instability (ROC of 0.82±0.03, sensitivity of 0.57±0.07 when the specificity was targeted at 0.90; the alert rate ratio of unstable to stable patients was 7.62). We conclude that these algorithms could form the basis for reliable predictive clinical alerts which identify patients likely to become hemodynamically unstable within the next few hours so that the clinicians can proactively manage these patients and provide necessary care.

Keywords: multiparameter monitoring, predictive alerts, acute hypotension, hemodynamic instability, ICU intervention

1. INTRODUCTION

Hemodynamic instability is considered one of the most critical events that require effective and prompt intervention in the intensive care unit (ICU). It is one of the major reasons for ICU recidivism which causes longer length of ICU stay [1]. It is most commonly associated with an abnormal or unstable blood pressure (BP), especially hypotension, or more broadly associated with inadequate global or regional perfusion. Inadequate perfusion may result in damage of vital organs. The heart and brain are particularly sensitive to perfusion compromise due to their high metabolic requirements and their limited ability to compensate for decreased oxygen and metabolic substrate

delivery. Compromised perfusion of these vital organs may quickly lead to life-threatening organ failure and death. Under normal circumstances, the autonomic nervous system provides fine control of cardiac output that is adequate to maintain body function and end organ perfusion in the face of normal stresses. However, in the ICU, clinicians often need to proactively intervene to augment the body's own feedback systems which may be inadequate to maintain perfusion of vital organs in the face of extreme stressors such as sepsis, hemorrhage or severe cardiac injury. In these instances, perfusion is often maintained through the judicious and timely use of intravenous (IV) fluids and blood transfusion for vascular filling and/or vasoactive and inotropic agents. The onset of significant hemodynamic instability is often an event that occurs at the end of a long series of subtle physiologic derangements in patients who are becoming unstable. These subtle derangements may not be readily noticeable within the torrent of physiologic data presented regularly to clinicians caring for patients in a modern ICU.

Current automated alert mechanisms typically generate an alert if one or more physiological parameters (e.g., mean BP) cross predetermined "normal" thresholds. Unfortunately, often by the time these alerts are triggered, the patient is already in an unstable state, making intervention much more difficult. Furthermore, these alerts are often subject to high false alert rates, adding to the information overload [2-4]. Early interventions in disease processes which often result in hemodynamic instability and end organ perfusion compromise such as sepsis have been widely advocated as a means to limit the morbidity and mortality [5,6]. Thus, reliable methods to predict which patients are most likely to become unable to maintain organ perfusion through their own autonomic control should help ICU caregivers focus on earlier interventions that would limit the impact of physiologic deterioration in the sickest patients.

We have previously reported a method to reduce the false alert rate by identifying artifacts using multiple channels of monitoring data [7]. We have also developed a rule-based algorithm to predict hemodynamic instability using *hourly* electronic charting data [8] and logistic regression models using *minute-by-minute* vital signs from electronic patient monitors [9].

In this work, we report on hemodynamic instability prediction with *minute-by-minute* segments of continuous monitoring data in, for example, 2 hours. The purpose of the algorithms we developed based on trend (minute-by-minute) data of vital signs was not to replace the existing alerting systems in the current ICUs, which catch hemodynamic instability events when they occur. Rather, the purpose of these algorithms was to predict these events ahead of time so that the clinicians can proactively manage these patients and reduce the number of hemodynamic instability events, leading to improved patient care and outcome. It is important to develop algorithms that use features that are robust to missing data, a common reality of the ICU data. For example, since spectral analysis using Fourier Transformation (FT) is not robust to missing data, it cannot be used to extract features in this case, although it has been shown to be very important in heart rate variability analysis [11]. Using a window (2-hour segment) makes the algorithm less sensitive to natural variations of the measured parameters and noisy artifacts. In addition, it becomes possible to extract useful pattern identifying instability from multiple data points. Therefore, a better

classification performance might be achieved using monitoring data of higher resolution.

In this work, hemodynamic instability was based on clinical interventions of vasopressors (more details in METHODS Section 2.1). In other words, prediction of hemodynamic instability was equivalent to prediction of clinical interventions presumably due to hemodynamic instability. Our hypothesis was that by combining information from several continuously monitored vital signs (i.e., HR from body surface ECG signals and arterial BP), more timely and accurate alerts can be generated than traditional alerts that are triggered when a single data stream crosses a threshold. In order to ensure that the alerts are accurate (low false alert rate) and timely, they were developed using a high targeted specificity and using data two hours before the patient's deteriorating condition was diagnosed and an intervention was executed (i.e., the intervention occurs two hours after the end of the 2-hour window of data, illustrated in the top panel of Figure 4).

The study was approved by the Institutional Review Boards of Beth Israel Deaconess Medical Center (Boston, MA) and the Massachusetts Institute of Technology (Cambridge, MA) during collaboration with Philips Healthcare (Andover, MA). Requirement for individual patient consent was waived because the study did not impact clinical care and all protected health information (PHI) was de-identified.

2. METHODS
2.1. Data collection and reference data sets
The reference data sets were based on retrospective data from MIMIC (Multiparameter Intelligent Monitoring in Intensive Care) II, which is a database of ICU patients created by a partnership between Philips Healthcare (Andover, MA), Massachusetts Institute of Technology (Cambridge, MA) and Beth Israel Deaconess Medical Center (Boston, MA) [10]. The version of the MIMIC II database that we used contained data of 12,695 medical, surgical, post-cardiac surgery, and cardiac ICU adult patients recorded between 2001 and 2005. A clinical information system, CareVue 9000 (Philips Healthcare, Andover, MA), was utilized to acquire laboratory results, IV medications, nurses' notes, and nurse validated monitoring data such as HR and BP on an hourly basis. Besides the CareVue data, the MIMIC II database also includes waveform data and minute-by-minute trend data saved directly from the patient monitors (IntelliVue MP-70 Patient Monitor, Philips Healthcare, Andover MA) for 1875 patients, from which our database was derived. The trend data include physiological data such as HR, various invasive BP measurements, oxygenation saturation (SpO2), and respiration rate (RR). Patient records with trend data for HR from ECG and arterial BP (systolic, diastolic and mean) were chosen for further analysis.

We developed our reference data sets based on whether a patient had therapeutic interventions indicative of hemodynamic instability. In particular, those patients who received vasopressors (dopamine, levophed, neosynephrine, epinephrine) were defined as *unstable*. In the patient population that did not qualify to be unstable, patients were defined as *nonstable* if they received at least one of the following: (1) medications such as lidocaine, nitroglycerine, labetolol, esmolol, nitroprusside, amiodarone, lasix, milrinone, (2) at least 1500cc IV-fluids in one hour, (3) at least 750cc of packed red

blood cells in 24 hours. The patients who were neither unstable nor nonstable were defined as *stable*. However, patients who had data gaps of more than 2 hours immediately after receiving vasopressors were excluded because they might be out of the ICU for surgery, and the vasopressor might have been administered for reasons other than hemodynamic instability. We utilized the occurrence of a "qualifying" clinical intervention to define hemodynamic instability as described above, rather than the simple occurrence of a change in vital signs, so that the physicians' clinical judgment on the ICU patients at the bedside could be incorporated in the definition. Although the treatment end-points might differ for each disease-state group, we aimed at identifying those with physiologic evidence of deterioration without regard to the treatment goals normally prescribed for a particular subset of patients in our population. Since hypotension requiring the use of vasopressor medication is a common final signal of hemodynamic instability among all of the potential subsets of patients included in our very heterogeneous population, we have utilized this as the standard for evidence of hemodynamic instability.

There were 66 stable patients and 104 unstable patients that met our criteria, and 83 non-stable patients that did not meet our instability criteria. There were 128 points in each two-hour segment, and we required no more than 5 points missing in each segment. For each stable patient, up to 10 two-hour segments of minute-by-minute trend data since admission were selected for a total of 505 segments. For each of the 104 unstable patients, a two-hour segment 2 hours before the first onset of a vasopressor medication was selected. Twelve additional unstable segments were identified for which the vasopressor therapy was reinitiated after a period of at least 24 hours without vasopressors (i.e., the segment began at least 24 hours after therapy stopped and ended 2 hours before another therapy was initiated). These 116 segments constituted the reference data sets for unstable patients. No "stable" segments were extracted from unstable patients in order to establish a clear-cut case scenario, since the underlying physiological state might have been changed well in advance when the clinicians initiated critical intervention for hemodynamic instability, and it would be unclear when an unstable patient was in a true stable physiological state.

2.2. Feature creation

Based on the *original* set of four physiological parameters – ECG heart rate (HR), systolic arterial BP (SAP), diastolic arterial BP (DAP), and mean arterial BP (MAP) - we developed several *derived* parameters. The intra-arterial BP data were collected directly from arterial line. The derived parameters include:

- abs_dHR: the absolute successive difference of the minute-by-minute HR data, a measure of physiological HR variation which is an indicator of many cardiovascular diseases [11].
- HR_div_BP: HR to BP ratio, including HR_div_SAP and HR_div_MAP, also known as the shock index, which may be an early indicator of cardiogenic, hypovolemic, and septic shock [12].
- HR_prod_BP: Heart rate pressure product, including HR_prod_SAP and HR_prod_MAP, associated with cardiac oxygen consumption [13].
- ECO: Estimated Cardiac Output, defined as k_{co}*HR*(SAP-DAP)/MAP, where

k_{co} is a constant proportionality related to the arterial compliance. This cardiac output model is a modified version of the Lijesstrand model (cardiac output ~ k*HR*(SBP-DBP)/(SBP+DBP)), with (SBP+DBP) replaced by MBP (mean blood pressure) [14,15]. This modified model was found to demonstrate the best performance [14]. In this paper, we set k_{co} as 1, and therefore, ECO is a signal that is proportional to the actual cardiac output which quantifies the mechanical work of the heart [14,15].

- Slope: calculated by fitting a linear least-square regression line to a moving window (4-, 8-, 16-, and 32-minute) for HR, SAP, MAP, and ECO, respectively.
- Slope projection: the projected value for HR, SAP, and MAP based on the slope for a moving window (4-, 8, 16-, and 32-minute). For example, the projected value of SAP using an 8-minute window is calculated by fitting a regression line to 8 minutes of data and then extrapolating it 8 minutes into the future. Table 1 summarizes the initial and derived physiological parameters used in this paper.

Table 1. Original and derived physiological parameters

Initial parameters	Derived parameters
- Heart Rate (HR) - Systolic Arterial Blood Pressure (SAP) - Mean Arterial blood pressure (MAP) - Diastolic Arterial Blood Pressure (DAP)	- abs_dHR (t)= |HR(t+1)-HR(t)| - HR_div_BP (t) = HR(t)/(BP(t), where BP can be SAP and MAP - HR_prod_BP (t)= HR(t)*BP(t), where BP can be SAP or MAP - ECO(t) = k_{co}*HR(t)*[SAP(t)-DAP(t)]/MAP(t), where k_{co} is a proportionality related to arterial compliance. - Slope (t): the linear regression slope of HR and BP at time t calculated for a time window (of, e.g., 4, 8, 16 and 32 minutes) from each 2-hour segment. - Slope projection (t): projected HR or BP values n minutes past current point at time t based on the slope of regression line for the last n minutes, where n can be 4, 8, 16, and 32 minutes.

t: time index.

To eliminate noise/outliers of each initial physiological parameter, we used a modified statistical outlier detector based on the deviation to the median and inter-quartile range values in a local moving window [16]. We then collected simple statistical measures

based on filtered parameters. The measures included extremes (i.e., minimum and maximum), moments (mean, standard deviation, and skewness), percentiles (5th, 10th, 50th, 90th, and 95th percentiles), and inter-percentile ranges (IPR, defined as the difference between two percentile values such as 95th and 5th percentiles). These measures are simple and easy to implement, and provide a snapshot of how each parameter is different for stable v.s. unstable patients. We refer to these measures as "features". Overall, we developed 220 features.

2.3. Feature selection and classification

We began feature selection by developing a univariable logistic regression model for each of the 220 features serving as predictors. Logistic regression models are one class of generalized linear models that assume $prob\{Y=1|X\}=1/[1+exp(-B_0-XB)]$ where Y is a binary outcome variable (0 or 1), X is a vector of the predictor variables, B_0 is the intercept, B is a vector of regression coefficients obtained from maximum likelihood estimation, and the left side of the equation gives the probability of Y =1 for a given X. In this paper, Y was interpreted as a diagnosis of instability. The probability calculated from the logistic regression model was called the instability index. Given a set of predictors (features), the instability index could be calculated. By setting all possible thresholds of instability index, a series of sensitivity and specificity, and therefore receiver operating characteristic (ROC) curve area could then be calculated. If the instability index was greater than the threshold, the 2-hour segment was classified as "unstable".

The study population is 505 two-hour segments from 66 stable patients and 116 two-hour segments from 104 unstable patients. The outcome variable, the probability of becoming unstable two hours later, is interpreted as an instability index. Wald's χ^2 test was used to test the significance of each coefficient in the model. In order to correct the impact of repeated measure on the artificial inflation of the sample size and therefore artificial increase in power in logistic regression models, we used the Huber-White method to adjust the variance-covariance matrix of a fit from maximum likelihood estimation. The p-values reported in the paper were from this robust variance-covariance matrix estimation. Only those features that were statistically significant (defined as $p < 0.05$) were considered. Only one feature from each physiological parameter (original or derived) was selected, unless there were two features that were both highly significant and weakly correlated (correlation coefficient < 0.2). There were 13 features that met our first group of criteria (shown in Table 2). Matlab 7.8 (The Mathworks, Inc., Natick, MA) was used for statistical analysis.

Table 2. Classification performance results using univariable logistic regression modeling ($p < 0.01$)

Feature	ROC	Sen1	Spc1	ss1	Sen2	ss1_spc90
p90HR_div_MAP	0.78 (0.73 0.82)	0.59 (0.49 0.68)	0.85 (0.82 0.88)	0.43 (0.31 0.56)	0.43 (0.34 0.53)	0.33 (0.21 0.45)
p90HR_div_SAP	0.77 (0.72 0.81)	0.69 (0.60 0.77)	0.73 (0.69 0.77)	0.42 (0.29 0.54)	0.40 (0.31 0.49)	0.30 (0.18 0.42)
p10SAP_slope_proj_8min	0.76 (0.71 0.81)	0.70 (0.61, 0.78)	0.71 (0.67 0.75)	0.41 (0.28 0.53)	0.39 (0.30 0.48)	0.29 (0.17 0.41)
p5SAP	0.75 (0.70 0.80)	0.72 (0.63 0.80)	0.66 (0.62 0.70)	0.39 (0.25 0.50)	0.41 (0.32 0.51)	0.31 (0.19 0.43)
p5MAP	0.74 (0.69 0.79)	0.71 (0.62 0.79)	0.72 (0.68 0.76)	0.43 (0.30 0.55)	0.34 (0.25 0.43)	0.25 (0.12 0.35)
stdHR_div_SAP	0.72 (0.66 0.78)	0.76 (0.67 0.83)	0.65 (0.61 0.69)	0.40 (0.28 0.52)	0.47 (0.38 0.57)	0.37 (0.25 0.49)
stdHR_div_MAP	0.71 (0.64 0.76)	0.66 (0.57 0.75)	0.72 (0.68 0.76)	0.38 (0.25 0.51)	0.35 (0.27 0.45)	0.25 (0.14 0.37)
p90_10_ECO	0.66 (0.60 0.71)	0.64 (0.54 0.73)	0.64 (0.60 0.68)	0.28 (0.14 0.41)	0.18 (0.12 0.26)	0.082 (0.01 0.19)
p50HR	0.66 (0.60 0.71)	0.60 (0.51 0.69)	0.68 (0.64 0.72)	0.28 (0.15 0.41)	0.23 (0.16 0.32)	0.13 (0.03 0.24)
p50_abs_dHR	0.66 (0.60 0.71)	0.60 (0.51 0.69)	0.67 (0.63 0.71)	0.27 (0.14 0.40)	0.20 (0.13 0.28)	0.14 (0.00 0.21)
p10HR_slope_4min	0.65 (0.59 0.70)	0.59 (0.49 0.68)	0.68 (0.64 0.72)	0.26 (0.13 0.4)	0.20 (0.13 0.28)	0.099 (0.00 0.21)
p10ECO_slope_16min	0.63 (0.58 0.69)	0.66 (0.57 0.75)	0.59 (0.55 0.63)	0.26 (0.12 0.38)	0.24 (0.17 0.33)	0.14 (0.04 0.25)
p95_5_HR_prod_SAP	0.57 (0.50 0.63)	0.34 (0.25 0.43)	0.82 (0.78 0.85)	0.15 (0.03 0.28)	0.22 (0.15 0.31)	0.12 (0.02 0.23)

95% confidence interval data are parenthesized "( )". ROC: Area under receiver-operating-characteristic curve; ss1: maximum (sensitivity+specificity-1); Sen1, Spc1: sensitivity and specificity values when ss1 is maximized over the whole range; Sen2, ss1_spc90: sensitivity

and ss1 values when specificity $\geq 90\%$. p50HR = 50th percentile of HR; p90HR_div_SAP = p90(HR/SAP); p90HR_div_MAP = p90(HR/MAP); p95_5_HR_prod_SAP = p95(HR*SAP) - p5(HR*SAP); p90_10_ECO = p90 (ECO) - p10(ECO); p10ECO_slope_16min = 10th percentile of ECO slope of moving 16-minute window; p50_abs_dHR = median of abs[HR(n+1)-HR(n)], n=1, 2,…N-1 (N is segment length); p10HR_slope_4min = 10th percentile of HR slope of moving 4-minute window; stdHR_div_SAP = standard deviation of (HR/SAP); stdHR_div_MAP = standard deviation of (HR/MAP); p5SAP, p5MAP = 5th percentile of SAP and MAP; p10SAP_slope_proj_8min = 10th percentile of SAP projected 8 minutes past current point based on slope of regression line for last 8 minutes.

In order to assess the correlation structure of these selected features, a hierarchical tree was constructed (Figure 1) using the R function varclus from package DESIGN to draw the dendrogram depicting the clusters, and chose the matrix of squared Spearman rank correlation coefficients as the similarity matrix (R, http://www.r-project.org/). The Spearman correlation matrix was chosen due to some obviously non-Gaussian variables such as p50_abs_dHR. One feature was selected from each uncorrelated sub-tree as candidate feature based on their ROC area values. The best combination of four features based on ROC area after testing on all possible combinations of the features was p90HR_div_MAP, p50_abs_dHR, p95_5_HR_prod_SAP, and p5SAP, where p90HR_div_MAP was from sub-tree p50HR, p90HR_div_SAP and p90HR_div_MAP; p50_abs_dHR was from sub-tree p50_abs_dHR and p10HR_slope_4min; p95_5_HR_prod_SAP was from sub-tree p95_5_HR_prod_SAP, p90_10_ECO, and p10ECO_slope_16min; p5SAP was from sub-tree stdHR_div_SAP, stdHR_div_MAP, p5MAP, p5SAP, and p10SAP_slope_proj_8min.

We also compared performance on various subsets of the selected features by multivariable logistic regression models. By setting different instability index thresholds, one can adjust the tradeoff between sensitivity and specificity and maximize their summation. We define ss1 = sensitivity + specificity-1, also known as Youden's index [17], as an indicator of classification accuracy. In addition, in order to maintain a low false alert rate, we set a threshold for feature values so that the targeted specificity was 0.90, and computed the corresponding sensitivity.

In order to show the advantages of our predictive algorithms, we compared the classification results using conventional alerts for physiological parameters outside of the normal ranges. The current clinical practice uses single systolic or mean blood pressure to alarm for potential hemodynamic instability. However, there are no universal thresholds for blood pressure levels. According to the guideline for septic patients [6], for example, the rule is systolic pressure < 90 mmHg or mean pressure < 65 mmHg. Therefore, we tested the classification performance using these two rules.
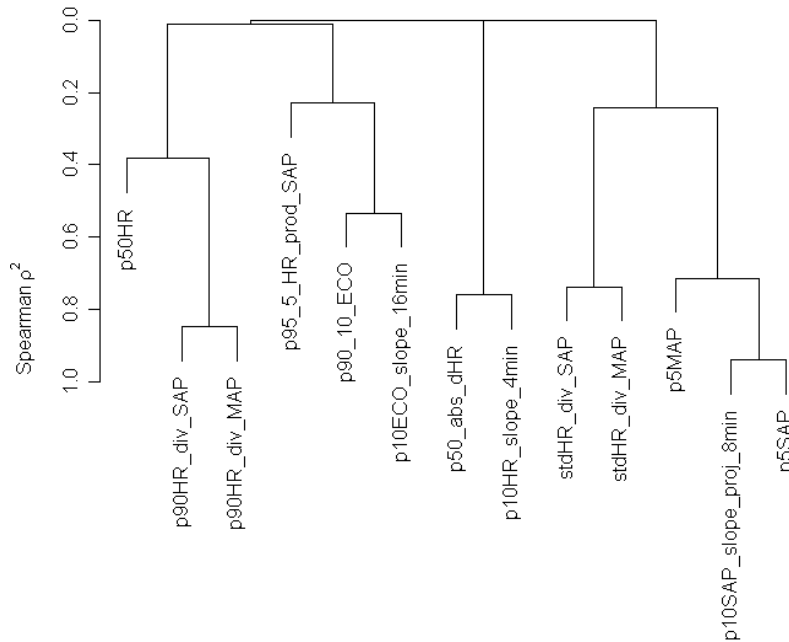
Figure 1.   Spearman correlation ρ^2 of the selected features. The features can be partitioned into four uncorrelated groups. One feature is selected from each group as a candidate feature.

2.4. Impact of prediction period on sensitivity

We defined "prediction period" as the length of time from the *end* of an unstable segment to the time of the *start* of the critical intervention. In the reference data, we used a prediction period of an arbitrarily 2 hours; *i.e.*, the end of the 2-hour segments were 2 hours before intervention (and the beginning of the 2-hour segments were 4 hours before intervention).

The prediction period of more than one hour was chosen because the intervention time recorded by the clinicians was often quite inaccurate, e.g., off by 30 minutes. On the other hand, if a prediction period was too long, the underlying physiological state might change. Therefore, a prediction period of two hours was chosen as falling in between these two extremes. Nevertheless, in order to assess how well our algorithms predicted deterioration as the prediction period was lengthened, we did an analysis for various length of prediction periods ranging from 0 to 12 hours (see RESULTS Section 3.2).

2.5. Impact of completeness of record on sensitivity and specificity

In real monitoring data, missing data is a common issue. It is important to develop algorithms to be robust enough to tolerate missing data when using a segment of data rather than single data points. The duration of missing data could range from one or a

few consecutive points (a few minutes) to any extended length (e.g., hours, days). We assessed the impact of completeness of record on classification performance for both stable and unstable patients. We compared the sensitivity and specificity when the required record completeness was varied over a range of 25%, 30%, 33.3%, 40%, 50%, 60%, 75%, 80%, 90%, and 95%. As a special case, the data was downsampled to 5-minute by 5-minute from minute-by-minute using median value of each 5-minute data (typically the median value of every 5 points if there was no missing data), and the impact of performance was also assessed.

2.6. Validation

We validated our regression models using bootstrapping. *Bootstrapping* is a name generically applied to statistical resampling schemes that allow uncertainty in the data to be assessed from the data themselves. The basic idea is, given n samples ($y1$, $y2$,…, yn) of a random variable $\mathbf{Y}$, which has an unknown cumulative distribution $F(y) = \text{Prob}\{\mathbf{Y} \leq y\}$, to compute the statistic of interest and to assess how the statistic behaves over B repetitions of sampling with replacement. As an estimate of $F(y)$, the empirical cumulative distribution $Fn(y)$ can be estimated from repetitive sampling with replacement from the n observed data when the number of repetition, B, is large enough. Therefore, the uncertainty or accuracy of the statistic of interest can be estimated empirically using confidence intervals, standard errors (SE), *etc*. We used $B = 100$ repetitions of the bootstrap validation procedure for each model. The classification performance was assessed by both ROC and contingency analysis. The validation results were reported as mean ± standard error (SE).

In addition, we validated our algorithm by relaxing the data selection criteria to include segments: (1) with up to 50% of missing data instead of 4%; (2) for the whole length of stay, instead of the first 10 segments only for stable patients, in the reference dataset; (3) with 120 points in each two-hour segment in the validation dataset, instead of 128 points in the reference dataset, in order to reflect the 2-hour in real monitoring practice. However, since only a few unstable segments were gained after relaxing the completeness from 96% to 50%, the data expansion was not sufficient to serve as a validation dataset for unstable patients. Therefore, we only validated our algorithms on the expanded dataset for stable patients.

2.7. Alert rate

High false alarm rates are a major concern in the current ICU clinical settings. In order to assess the additional workload our alert algorithms might add to clinicians in the clinical settings, as well as to assess the effectiveness of the prediction of deterioration in real practice, we calculated and compared the alert rates for both stable and unstable patients.

In many cases for classification algorithms, a positive predictive value (PPV) is reported to assess the percentage or probability of true positives among all positives. However, in real-time monitoring cases, rather than those diagnostic screening tests, PPV is dependent upon the relative length of time the stable patients are exposed to the alerting algorithms compared to the unstable patients. It should also be stressed that the relative exposure times are not only dependent upon the size of the two populations, but also upon the average length of time the two groups are exposed to the algorithm. For

instance, an exposure ratio of 5 could be due to a situation where there are an equal number of stable and unstable patients, but the algorithm runs 5 times longer on the stable patients (their complete stay) than on the unstable patients (episodes prior to interventions). Therefore, in this section we present the relative alert rates that are not as much affected by the ratio of the exposure time of stable and unstable patients.

Rule firings were defined as the time when the instability index of a segment exceeded the preset threshold. Because the data were largely correlated, the firing tended to occur consecutively and unnecessarily frequently. In order to address this problem, we defined an *alert* based on a *refractory period* as the following: after the first rule firing, new rule firings were suppressed until there were no more rule firings for a certain period of time (*e.g.*, 2 hours). In other words, consecutive firings were counted as one alert, and a new alert was counted only when there was a gap of at least a period of time (*e.g.*, 2 hours) between two firings. This specified period of silent time was defined as the refractory period. A two-hour refractory period was selected to ensure no consecutive firings from overlapping moving windows.

We calculated *alert rates* in three different ways. One way was to calculate an *aggregate* alert rate by dividing the total number of alerts of all patients by the total monitoring time (Alert rate 1). Suppose F_i is the number of alerts for patient i, and T_i is the monitoring time for the same patient. Alert rate 1 (AR1) is defined as:

$$AR1 = (\Sigma F_i)/(\Sigma T_i) \tag{1}$$

The second way (Alert rate 2, AR2) was to *average* the individual alert rates for each patient:

$$AR2 = 1/n \, \Sigma(F_i/T_i) \tag{2}$$

The third way is to use a Poisson regression method described below. Since the events occurred at a particular rate within a particular amount of time, an appropriate way to obtain a mean alert rate is to use Poisson regression models for predicting the expected value of the count given a time frame (e.g., one day); i.e., $\log(E[Y]) = a+bX$, where Y is the event count, E[Y] is the expected value of Y, and X is a vector of predictor variables. Therefore, we developed a third way of calculating alert rate using a Poisson regression model (using length of monitoring time as a predictor variable and number of alerts as dependent variable). For stable patients, the distribution of monitoring time was skewed to the right, resulting from long ICU stays of a few patients, so we took the logarithm of monitoring time as a predictor variable. For unstable patients, however, since the monitoring time was restricted to one to six hours only, the distribution is less skewed, and taking the logarithm of monitoring time as a predictor variable was not necessary. By setting the monitoring time as one day, the expected count of alerts per patient day can be determined. Alert rate 3 (AR3) is thus defined as:

$$\log(E[AR3]) = a + b \, X \tag{3}$$

where X is monitoring time, a is the intercept, and b is the coefficient of X.

When calculating alert rates, instead of using non-overlapping windows, we used moving 2-hour windows shifted in 15-minute increments to mimic a possible real time clinical implementation of our algorithms. Before calculating the alert rates, the *monitoring time* for each patient needs to be computed. In general, the time when there was no data should be excluded from the calculation of monitoring time in order to avoid an artificial inflation of monitoring time and therefore an artificial low alert rate. This is not straightforward because of the presence of missing data and the nature of overlapping moving windows. To accurately compute monitoring time, we set the *center* of the 2-hour operating window as the working point, which starts at the beginning of each patient's HR and arterial blood pressure (ABP) data. As the working point moves ahead in 15-minute increments, if the 2-hour segment meets a minimum requirement of completeness (e.g., 50%), 15 minutes are added to the monitoring time. If the 2-hour segment has less than 50% of data, the 15-minute time is not added. For example, if we have 60-minute continuous data, ideally, the monitoring time is 1 hour. If we need to move 4 times for the working point to be out of the 1-hour region, and each time 15-minute time is counted, that makes exactly 60 minutes (4×15 minutes).

3. RESULTS
3.1. Classification
Clinicians consider increased HR and decreased BP to be early signs of patient deterioration. Thus we developed statistical HR and BP features including the slope and hypothesized that they would be useful in predicting hemodynamic instability. Figure 2 shows the boxplots of the four sample features calculated for stable and unstable patients. Not surprisingly, unstable patients had lower BP and higher shock index. Also, unstable patients had lower physiological HR variation, which was consistent with other studies demonstrating that low beat-to-beat HR variability was usually a sign of physiological derangement [11]. Interestingly, we also found that unstable patients had slightly, but significantly higher rate pressure product, which might represent attempted physiological compensation — as BP decreases, HR often increases to compensate [18].

We also investigated the performance for multiple features. The results for the strongest performers for 2, 3 and 4 features are shown in Table 3. Using more than 4 features did not improve classification significantly (data not shown). Figure 3 shows the ROC curve when using the 4 features in Table 3 as predictors. Figure 4 shows an example of how the instability index changes before and after initiation of the intervention for an unstable patient.
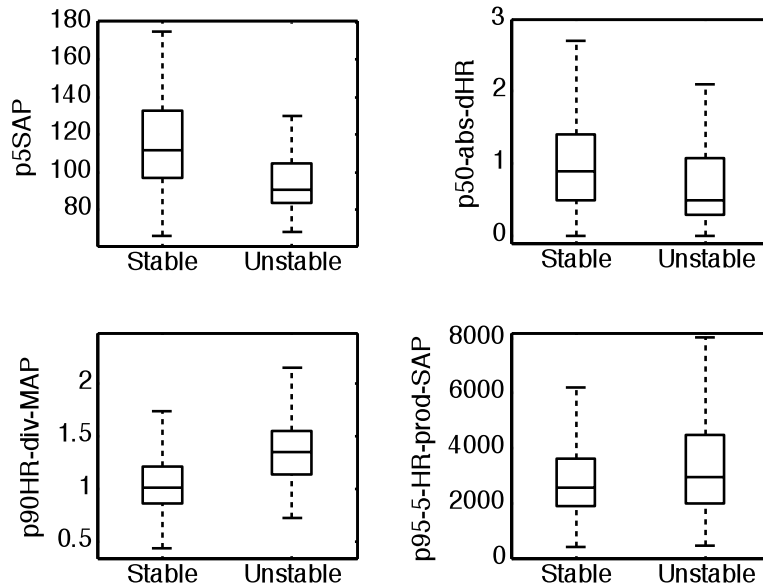
Figure 2. Boxplots of the four selected features for stable and unstable patients, respectively. Unstable patients had significantly lower BP and HR variation, higher shock index, and rate pressure product (Wald's χ^2 test, $p < 0.01$). p5SAP = 5th percentile of SAP; p50_abs_dHR = median of abs[HR(n+1)-HR(n)], n=1, 2,…N-1 (N is segment length); p90HR_div_MAP = p90(HR/MAP); p95_5_HR_prod_SAP = p95(HR*SAP) - p5(HR*SAP).

Table 3. Classification performance results using multivariable logistic regression modeling ($p < 0.05$)

Feature	ROC	Sen1	Spc1	ss1	Sen2	ss1_spc90
p90HR_div_MAP p50_abs_dHR	0.79 (0.75 0.84)	0.66 (0.57 0.75)	0.80 (0.76 0.83)	0.46 (0.33 0.58)	0.49 (0.40 0.59)	0.39 (0.27 0.51)
p90HR_div_MAP p50_abs_dHR p5SAP	0.81 (0.76 0.85)	0.70 (0.61 0.78)	0.81 (0.77 0.84)	0.51 (0.38 0.62)	0.53 (0.43 0.62)	0.43 (0.30 0.54)
p90HR_div_MAP p50_abs_dHR p5SAP p95_5_HR_prod_SAP	0.82 (0.77 0.86)	0.74 (0.65 0.82)	0.80 (0.76 0.83)	0.54 (0.41 0.65)	0.55 (0.46 0.64)	0.45 (0.33 0.57)

Note: Keys are the same as Table 2.
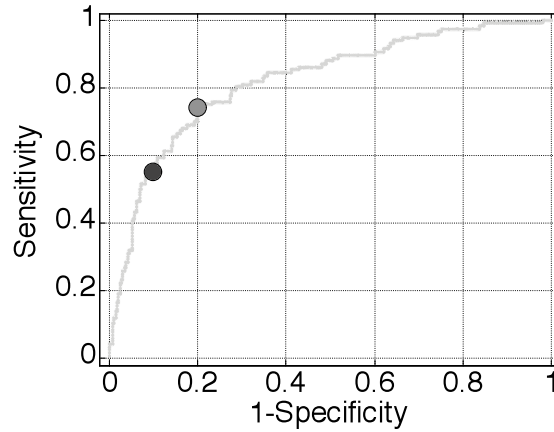
Figure 3.   ROC curve based on the same four features as in Figure 2 for
                 classification when the reference dataset was used. The upper grey circle
                 marks the optimal trade-off of sensitivity (0.74) and specificity (0.80)
                 with ss1=0.54. The lower dark circle marks the sensitivity (0.55) when
                 the specificity was targeted at 0.90.

We compared the performance of our algorithm with conventional alerts for
physiological parameters outside the normal ranges. Since there are no universal
thresholds for blood pressure levels for defining hemodynamic instability, as an
example, according to the guideline for septic patients [6], the rule is systolic pressure
< 90 mmHg or mean pressure < 65 mmHg. The classification results using these two
thresholds on our reference dataset are shown in Table 4. Compared to the classification
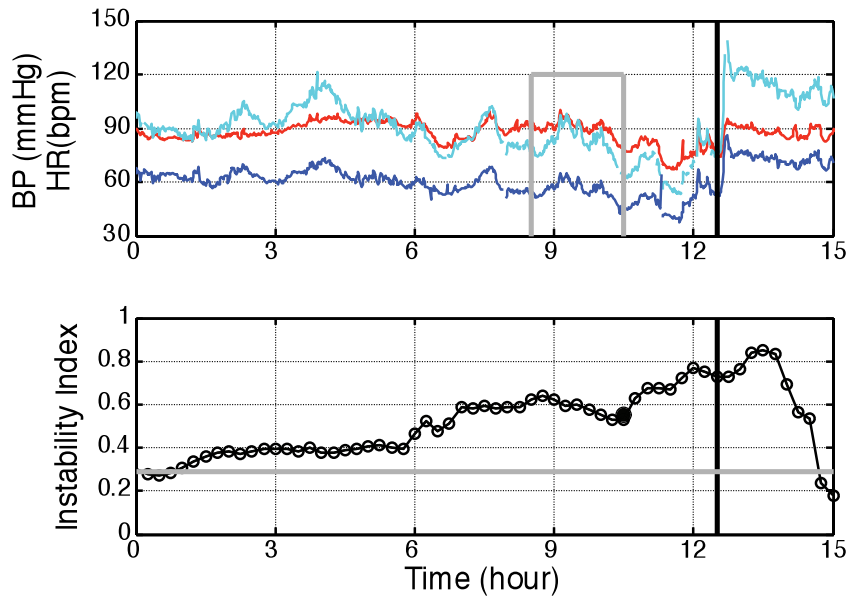performance from our predictive model, our results are better.

Figure 4.   An example showing the trend data (15 hours) from an unstable patient and the instability index based on logistic regression. For both panels, the black vertical line marks the beginning of the intervention. The time zero is an arbitrary starting point of the ICU duration simply for demonstration purpose, not the actual ICU or hospital admission time. The top panel shows the trend data for HR (red), SAP (cyan), and MAP (blue), respectively. The grey box in the top panel marks the moving 2-hour segment at the time used in reference dataset – two hours before the intervention. The bottom panel shows the predicted instability index using a 2-hour segment shifted in 15-minute increments. The instability index calculated from this segment is marked by the solid circle in the bottom panel. The grey horizontal line in the bottom panel marks the instability index threshold (0.29) when the specificity is targeted at 0.90. In other words, at a threshold of instability index of 0.29 calculated from the logistic regression model, at least 90% of stable segments are classified as "stable" by the model. This figure shows the following: (a) An early detection 11 hours prior to intervention initiation; (b) An upward trend of the instability index before initiation of the intervention and a downward trend afterwards, but with a delay that reflects the size of the window.

Table 4. Comparison of conventional alerts and the present predictive algorithms

Algorithms	ROC	Sensitivity	Specificity	ss1
SAP < 90 mmHg	0.70 (0.65, 0.75)	0.59 (0.50, 0.68)	0.80 (0.77, 0.83)	0.39 (0.27, 0.51)
MAP < 65 mmHg	0.72 (0.67, 0.76)	0.75 (0.66, 0.82)	0.69 (0.65, 0.72)	0.44 (0.31, 0.54)
Our algorithms	0.82 (0.77, 0.86)	0.74 (0.65, 0.82)	0.80 (0.76, 0.83)	0.54 (0.41, 0.65)

95% confidence interval data are parenthesized " ( )".

3.2. Impact of prediction period on sensitivity

Figure 5 shows that as the moving window approaches the time of deterioration (time 0), the instability index (top panel) increases, so does the proportion of unstable segments (bottom panel) that are correctly classified as unstable (*i.e.,* sensitivity) by our algorithm (when specificity was targeted at 90%). It drops quickly when the prediction period is longer than three hours. For example, sensitivity drops from 0.58 for *one* hour before intervention to 0.46 for *five* hours before intervention. Such decrease in prediction power was consistent with the finding shown in hemodynamic instability advisory using hourly clinical electronic charting data [8].

3.3. Impact of completeness of record on sensitivity and specificity

The impact of completeness of record on the sensitivity and specificity was assessed using the following two different instability index thresholds obtained from logistic regression model and the reference dataset: (1) 0.20, when the maximum ss1 was reached, and (2) 0.29, when the specificity was targeted at 90% (i.e. when a threshold of instability index of more than 0.29 was chosen, at least 90% of stable segments were classified as "stable" by the model). The data for the whole length of stay for stable patients and 2 to 8 hours before intervention for unstable patients were used for analysis of impact of missing data. As shown in Table 5, when the required record completeness varied from 95% to 25%, the changes in sensitivity and specificity were small ($\leq 2\%$), and there are a gain of 3449 (36.1%) 2-hour overlapping segments (shifted in 15-minute increments) for stable patients, and a gain of 548 (24.2%) for unstable patients. Note that the baseline specificity (95% completeness) was lower than shown earlier (Table 3) as a much earlier prediction period (2 to 8 hours, vs. 2 hours, before intervention) was used.
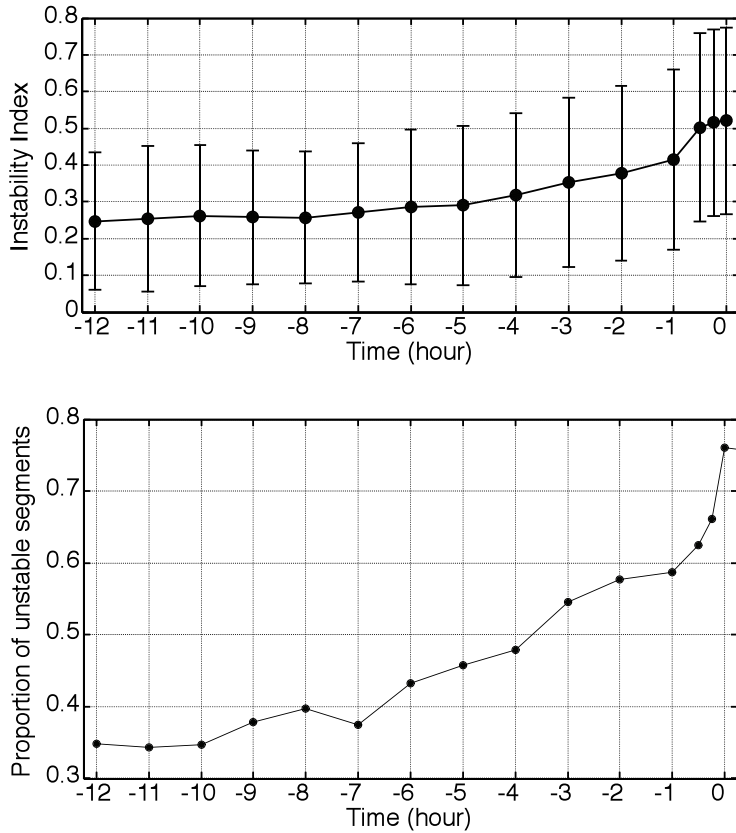
Figure 5. Instability index (top panel, mean±std) and probability density function of unstable segments correctly classified (viz. sensitivity) (bottom panel) as a function of time to intervention (i.e., prediction period). The specificity was targeted at 90%; i.e., 90% of the stable segments are less than the threshold of instability index 0.29, when the sensitivity was calculated. It shows that both the instability index and the proportion of unstable segments increase, as the prediction period approaches the time to intervention. The mean instability index is above the threshold of 0.29 when the prediction period is less than 5 hours, but not all unstable segments are classified as unstable because of the large standard deviation. Note that the "-n" labels on x-axis, representing the time to intervention, are actually 1 hour bins of *n* to *n+1 hours before intervention*. The two data points that fall between hour -1 and 0 on the x-axis are 30 minute and 15 minutes before intervention, respectively.

Table 5. Impact of completeness of record on classification indexes

Required record completeness	Threshold = 0.20		Threshold = 0.29		Number of segments	
	Sensitivity	Specificity	Sensitivity	Specificity	Stable	Unstable
≥ 95%	60.5%	80.4%	45.9%	88.7%	9557	2260
≥ 25%	62.0%	80.7%	47.9%	89.3%	13006	2808

Table 6 shows the impact of downsample from minute-by-minute data to 5-minute-by-5-minute data on the classification performance based on logistic regression modeling using the same 4 features described in Table 2. The first two rows are results from logistic regression models developed using 1-minute and 5-minute data, respectively, with the 4 selected features. The third row represents the performance on 5-minute data when the model was developed from 1-minute data and the instability index thresholds were 0.20 (when the maximum ss1 was reached) and 0.29 (when the specificity was targeted at 90%), respectively. The performance on less complete dataset and 5-minute data was similar to that of 1-minute data, indicating that our algorithm may not be sensitive to data completeness or sampling rate.

Table 6. Impact of downsampling on performance

Dataset	ROC	Sen1	Spc1	ss1	Sen2	ss1_spc90
1-minute	0.82	0.74	0.80	0.54	0.55	0.45
5-minute	0.82	0.76	0.76	0.52	0.52	0.42
5-minute validation (using logistic regression model from 1-minute reference dataset)	0.82	0.66	0.81	0.47	0.53	0.42

Note: Keys are the same as Table 2.

3.4. Validation
We validated the logistic regression model via 100 repetitions of the bootstrap validation procedure using p90HR_div_MAP, p50_abs_dHR, p95_5_HR_prod_SAP, and p5SAP as predictors. The classification performance was assessed by both ROC and contingency analysis. The validation results were reported as mean ± standard error in Table 7. The tightness of the variance limit for these performance indexes indicates that they are reliable. The results of 10-fold cross validation were similar (not shown).

Table 7. Validation results by multivariate logistic regression modeling via bootstrapping 100 times (p <0.05)

ROC	Sen1	Spc1	ss1	Sen2	ss1_spc90
0.82±0.03	0.75±0.07	0.80±0.05	0.56±0.05	0.57±0.07	0.47±0.07

Note: Keys are the same as Table 2.

In addition, we validated the algorithms by relaxing the data selection criteria and creating additional 2-hour segments from the MIMIC II database, as shown in Table 8. We allowed up to 50% (compared to 4%) of data to be missing in each 2-hour segment, and for the whole length of stay instead of the first 10 2-hour segments only. We identified 78 stable patients (compared to 66) with 1585 stable segments (compared to 505). The performance validation results (Table 6) showed that when using the same instability index threshold as for the reference data, the specificity was about the same (0.81 compared to 0.80 when the maximum ss1 was reached in the reference dataset; 0.89 when specificity was targeted at 0.90), even though it is on a much larger dataset with less stringent data requirements. The instability index thresholds of 0.20 and 0.29 are the same as that for Table 5.

Table 8. Validation results by relaxing data selection criteria

Dataset	Number of stable patients	Number of stable segments	Specificity (threshold = 0.20)	Specificity (threshold = 0.29)
Reference	66	505	0.80	0.90
Validation	78	1585	0.81	0.89

The algorithm appeared robust to missing data and data from the same database. However, an external validation on a new dataset is needed to make a better assessment of the prediction performance of these features.

3.5. Alert rate

We calculated the alert rates in order to estimate the workload to the clinicians if they adopt our algorithms in their clinical practice, since high false alert rates and the consequent unnecessary attention are problems for the current ICU monitoring techniques. We used 2-hour segments shifted in 15-minute increments (referred to as moving window of 15-minute) and a refractory period of 2 hours in order to mimic the real monitoring situation. Since only $\geq$ 50% of completeness or 1 hour of data was required for each 2-hour segment, there were 13 more stable patients (n=79) and 12 more unstable patients (n=116) than the reference dataset (The impact of data completeness is addressed in Section 3.3). Table 9 compares the alert rates for stable and unstable patients using aggregate and mean individual rates, respectively.

Table 9. Alert rates for stable and unstable patients

	Stable	Unstable	Alert rate ratio: unstable/stable
Aggregate alert rate (per patient day)	0.63	3.20	5.08
Mean individual alert rate (per patient day)	0.95	3.48	3.66
Alert rate from Poisson regression (per patient day)	0.47	3.58	7.62
Total monitoring time (day)	132.4	28.3	-
Total number of alerts	84	92	-
Number of patients	79	116	-

Note that the mean individual alert rate ratio was considerably lower (3.66) than that of the aggregate (5.08). This was largely due to the difference in alert rates for stable patients (0.95 vs. 0.63). We found that the distribution of the alerts for stable patients was non-symmetrical, with a high density at 0 alerts (no alerts in 62.0% of all 15-minute moving windows), and a long tail toward a large number of alerts (with a maximum number of alerts of 10). The spectrum of the number of alerts for stable patients is shown in Table 10. In addition, 1 out of 79 stable patients had one alert, but if the length of stay for this patient was only 1.25 hours, the resulting alert rate for this patient was 19.2 per patient day. By just excluding this patient, the mean individual alert rate dropped to 0.72 from 0.95 per patient day. Mean individual alert rate appeared to be biased by both a non-Gaussian distribution and outliers (viz. patients with short lengths of stay) for the stable patients. Furthermore, for the mean individual alert rate calculation, a patient with no alerts for a long period of time (e.g., 5 days) counts equally with another patient with no alerts for a short period (e.g., 5 hours). Both of them have an alert rate of 0, and are therefore weighted the same, unlike in the aggregate alert rate calculation. The alert rates from the Poisson regression models, however, were not biased by non-Gaussian distribution of the number of the alerts. In addition, in order to suppress the bias caused by the outliers of individual length of stay, we took the logarithm of monitoring time for stable patients. Therefore, the alert rate from the Poisson regression is the best estimator of alert rate among the three estimators of alert rates.

Table 10. Spectrum of alerts for stable patients (n=79)

Number of alerts	0	1	2	3	4	5	6	7	10
Number of patients	49	13	5	1	6	1	2	1	1
Percentage of patients (%)	62.8	16.7	6.4	1.3	7.7	1.3	2.6	1.3	1.3

Given relative alert rates and exposure times, one can estimate the PPV as TP/(TP+FP), where TP is true positive, FP is false positive . For example, using the aggregate alert rates presented in this paper (Table 9) with an unstable-to-stable-alerts ratio of about 5:1 and assuming that the stable patients are exposed to the alerts 5 times longer than the unstable patients (in the aggregate), then the number of TP will be equal to that of FP, and therefore the PPV would be 0.5.

4. DISCUSSION

ICU clinicians often base clinical care decisions on multiple physiological parameters and their trends. Traditional alerting technologies use single parameter and single threshold to alarm clinicians for attention. This work suggests that there may be physiologic features that can be predictive of impending hemodynamic deterioration – and subsequent need for aggressive therapy – at least two hours prior to the onset of therapy. We found that hemodynamically unstable patients had lower BP, higher shock index, lower HR variability, and higher rate pressure product than stable patients. The latter three features are associated with physiological compensation mechanisms of low BP. For an optimal tradeoff of specificity and sensitivity, logistic regression achieved a specificity and sensitivity of 0.80 ± 0.07 and 0.75 ± 0.06, respectively. When the specificity was targeted at 0.90, the sensitivity was 0.57 ± 0.07. By targeting at a high specificity, the false alert rate decreased. The present algorithm appeared robust to missing data. The aggregate alert rate was 0.63 and 3.20 per patient day for stable and unstable patients, respectively, and the alert rate ratio of unstable vs. stable patients was 7.62.

4.1. Feature extraction

When the same single features were used, the ROC results of the present algorithms were similar to the results of one of the winners of the *Computers in Cardiology Challenge* 2009 [19], even though there were a number of differences between the two (for example, to predict well-defined and hand-picked cases of acute hypotensive events rather than clinical intervention of pressor, and different prediction windows of one hour vs. two hours). One of the strengths of our algorithms is that it enables using multiple features to predict patient deterioration, and the classification performance for multi-feature was better than single feature, e.g., sensitivity of 0.55 (4 features extracted from HR and BP, Table 3) compared to 0.43 (using best single feature of p90HR_div_MAP, Table 2) when specificity was targeted at 0.90.

In addition to the features described above, we explored a number of simple and complex features (e.g., from wavelet transform of different scales), but there was no significant improvement in performance. Saeed et al. introduced a predictive algorithm using symbolic representations of wavelet representations of hemodynamic time series and the MIMIC II database [14, 20]. The performance results of the present study were similar to theirs. However, given the limitations of our reference data set, it is premature to conclude that complex features such as ECG signal and beat-to-beat HRV have nothing to contribute. In particular, the features utilized in the present study were based on features chosen *a priori* that are likely to be physiologically meaningful. Robust "unsupervised" data mining techniques with MIMIC II may reveal hitherto

undiscovered features that may be predictive of one or more classes of hemodynamic instability.

4.2. Clinical relevance of hemodynamic control and reasons for false negatives and positives

Reduced BP and its downward trend are the main indicators of hemodynamic instability in clinical practice, consistent with our findings. Shock index, defined as the ratio of HR to BP, was the primary predictor of hemodynamic instability in our findings, but not widely used in clinical practice. Shock index was not correlated with SAP (correlation coefficient < 0.1). Low HR variability has been associated with various disease states [11], and we found that unstable patients tended to have lower physiological HR variation than stable patients, a result consistent with the findings of the studies that used HR and BP spectra to assess the autonomic cardiovascular regulation [21]. Rate pressure product was reported to be associated with mortality in trauma ICU patients [22]. This combination of features yielded a better classification performance than a single feature such as BP (ROC 0.82 vs. 0.75).

In order to improve our prediction algorithms, we attempted to determine the reasons for false negatives and false positives by examining the nurses' notes for additional information about patients, such as diagnosis, surgery, and medication. The following summarizes the reasons.

We identified two categories of reasons for *false negatives* where our algorithms failed to classify unstable segments as unstable. *First*, some patients simply deteriorated too fast (within 2-hour window) to be caught by our algorithms. *Second*, there was background information available to the clinician but not to the algorithms, and the clinicians intervened due to patients' conditions other than hemodynamic instability. These conditions were confirmed by reviewing the individual patient's medical records, and included the following: (a) Intra Aortic Balloon Pump (IABP) patients who were hemodynamically unstable by definition, but their physiological parameters of HR and BP may appear stable, and our data selection did not exclude them. Thus a number of false negatives were due to our inability to exclude IABP patients in our data selection. (b) Patients receiving a vasopressor after cardiac surgery to manage mild to severe hypotension, or to overcome the excessive effect of nitroprusside during high blood pressure control, rather than hemodynamic deterioration. (c) Head injury and spinal cord injury patients who receive neosynephrine and labetalol simultaneously to control BP and to ensure sufficient cerebral perfusion pressure (CPP). (d) Head injury and spinal cord injury patients with vasopressors as a therapeutic measure in order to maintain CPP above 70 mmHg in the phase of elevated intracranial pressure (ICP) by keeping the BP at elevated levels (e.g., mean BP > 80 mmHg or systolic BP > 95 mmHg) [23,24]. Thus, a number of false negatives (above cases (b) to (d)) were noted because patients received vasopressors for therapeutic purposes not related to hemodynamic instability. These patients were not hemodynamically unstable from a clinical perspective. For instance, the head injury patients in our study often received vasopressor as a therapeutic measure to maintain CPP in the setting of increased ICP. In these cases, the administration of vasopressor therapy did not coincide with true hemodynamic instability. However, since we used administration of vasopressors as the

criterion of instability (rationale of using this criterion is given in METHODS Section 2.1), these patients were labeled as unstable in our dataset, which confounded the present results.

The main reason for *false positives*, where our algorithm failed to classify stable patients as stable, was when a low BP (*e.g.*, SAP < 90 mmHg) was accompanied by a consistently high HR (*e.g.*, HR > 100 bpm), resulting in a high instability index, since one of the major predictors was the ratio of HR to BP (viz. p90_HR_div_BP). This happened in 7 out of 10 patients whose instability index well exceeded the threshold when targeting a specificity of 90% (a threshold of 0.50 vs. 0.29, see RESULTS Section 3.3) using our predictive algorithm. However, physicians sometimes missed these events either because they failed to notice this pattern indicating a need for intervention and the patients recovered later on, or they did not consider it to be serious enough to initiate an intervention that met our definition (viz. major vasopressor administration). Instead, the clinicians may have chosen a less aggressive treatment for the episode of hemodynamic instability that did not meet our absolute definitions for an intervention (e.g., an infusion of 100cc/hour of normal saline for several consecutive hours). Other reasons of false positives included positional or dampened waveform data arising from the arterial line collecting BP data [25]. In this case, the monitor continued to report artificially low BP values, but the clinicians at the bedside recognized the condition from the resulting waveform and chose not to initiate therapy knowing the arterial line data was falsely reporting low BP. Ideally, the BP data from the overdampened waveform should be removed from the database using waveform data. However, it has also been demonstrated that the dampened data could also be removed using arterial BP time series and the data usefulness could be potentially improved [26]. This will be one area for our future algorithm improvement.

Our next steps include improving instability prediction by combining clinical lab data with trend data, as well as validating the algorithms on new datasets.

5. CONCLUSIONS

Timely identification of patients who are likely to become hemodynamically unstable would enable earlier intervention which will limit organ injury associated with low perfusion events in the ICU. The algorithms presented in this work based on trend (minute-by-minute) data of vital signs could form the basis for reliable predictive clinical alerts which identify patients likely to become hemodynamically unstable so that the clinicians can proactively manage these patients and reduce the number of hemodynamic instability events, leading to improved patient care and outcome.

NOMENCLATURE

abs_dHR (t)	absolute difference between two consecutive heart rate data points, i.e., |HR(t+1)-HR(t)|, where t is the index of the current time, t+1 is the index of the next time
ECO (t)	estimated cardiac output, defined as k_{co}*HR(t)*[SAP(t)-DAP(t)]/MAP(t), where k_{co} is a proportionality related to arterial compliance
FP	false positive
HR_div_BP (t)	ratio of heart rate and blood pressure values at time t, where BP can be SAP and MAP
HR_prod_BP (t)	product of heart rate and blood pressure values at time t, where BP can be SAP or MAP
p10ECO_slope_16min	10th percentile of ECO slope of moving 16-minute window
p10HR_slope_4min	10th percentile of HR slope of moving 4-minute window
p10SAP_slope_proj_8min	10th percentile of SAP projected 8 minutes past current point based on slope of regression line for last 8 minutes
p50_abs_dHR	median of absolute difference of successive heart rate data
p50HR	50th percentile of heart rate values in each segment
p5SAP, p5MAP	5th percentile of SAP and MAP
p90_10_ECO	difference of 90th and 10th percentiles of ECO
p90HR_div_MAP	90th percentile of the ratio of heart rate to mean arterial blood pressure values
p90HR_div_SAP	90th percentile of the ratio of heart rate to systolic arterial blood pressure value
p95_5_HR_prod_SAP	difference between 95th and 5th percentiles of the product of heart rate and systolic arterial blood pressure
PPV	positive predictive value
Sen1, Spc1	sensitivity and specificity values when ss1 is maximized over the whole range
Sen2, ss1_spc90	sensitivity and ss1 values when specificity = 90%
Slope (t)	the linear regression slope of HR and BP at time t calculated for a time window (e.g., of 4-, 8-, 16- and 32-minute) from each 2-hour segment
Slope projection (t)	projected HR or BP values n minutes past current point at time t based on the slope of regression line for the last n minutes, where n can be 4-, 8-, 16-, and 32-minute
stdHR_div_MAP	standard deviation of the ratio of HR and MAP
stdHR_div_SAP	standard deviation of the ratio of HR and SAP
TP	true positive

REFERENCES

[1] Vohra, H.A., Goldsmith, I.R., et al, The predictors and outcome recidivism in cardiac ICUs., *European Journal of Cardio-Thoracic Surgery*, 2005, 27(3), 508-11.

[2] Tsien, C.L., Fackler, J.C., Poor prognosis for existing monitors in the intensive care unit, *Crit Care Med*, 1997, 25(4), 614-9.

[3] Chambrin, M.C., Alarms in the intensive care unit: how can the number of false alarms be reduced? *Crit Care*, 2001, 5(4), 184-8.

[4] Aboukhalil, A., Nielsen, L., Saeed, M., Mark, R. G. & Clifford, G. D. 2008 Reducing false alarm rates for critical arrhythmia using the arterial blood pressure waveform. J. Biomed. Inform. 41, 442-451.

[5] Casserly, B., Read, R., Levy M.M., Hemodynamic monitoring in sepsis, *Crit Care Clin*, 2009, 25(4), 803-23.

[6] Anonymous, Practice parameters for hemodynamic support of sepsis in adult patients in sepsis. Task Force of the American College of Critical Care Medicine, Society of Critical Care Medicine, *Crit Care Med*, 1999, 27(3), 639-60.

[7] Ali, W., Eshelman, L.J., Morphograms: exploiting correlation patterns to efficiently identify clinically significant events in intensive care units. In *Conf Proc 2004 IEEE Eng Med Biol Soc*, 2004, 1, 554-7.

[8] Eshelman L.J., Lee K.P., et al, Development and evaluation of predictive alerts for hemodynamic instability in ICU patients, in *AMIA Annu Symp Proc 2008*, 2008, 379-83.

[9] Cao, H., Eshelman, L.J., et al., Predicting ICU Hemodynamic Instability Using Continuous Multiparameter Trends. In *Conf Proc IEEE Eng Med Biol Soc.2008*, 2008, 3803-6.

[10] Saeed, M., Lieu, C., et al, MIMIC II: a massive temporal ICU patient database to support research in intelligent patient monitoring. In *Comput Cardiol*, 2002, 29, 641-4.

[11] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, heart rate variability: standards of measurement, physiological interpretation and clinical use, *Circulation*, 1996, 93(5), 1043-65.

[12] Rady, M.Y., Smithline, H.A., et al, A comparison of the shock index and conventional vital signs to identify acute, critical illness in the Emergency Department, *Ann Emerg Med*, 1994, 24(4), 685-90.

[13] Gobel, F.L., Norstrom, L.A., The rate-pressure product as an index of myocardial oxygen consumption during exercise in patients with angina pectoris, *Circulation*, 1978, 57(3), 549-56.

[14] Saeed, M., Temporal pattern recognition in multiparameter ICU data, Ph.D. dissertation, Massachusetts Institute of Technology, 2007.

[15] Sun J.X., Reisner A.T., et al., The cardiac output from blood pressure algorithms trial, *Crit Care Med*, 2009, 37(1), 72-80.

[16] Cao, H., Lake, D.E., et al, Toward quantitative fetal heart rate monitoring, *IEEE Trans Biomed Eng*, 2006, 53(1), 111-8.

[17] Hilden, J., Glasziou, P., Regret graphs, diagnostic uncertainty and Youden's index, *Statistics in Medicine*, 1996, 15, 969-86.

[18] Gross, B.D., Sacristán, E., et al, Supplemental Systemic Oxygen Support Using an Intestinal Intraluminal Membrane Oxygenator, *Artificial Organs*, 2000, 24(11), 864–9.

[19] Chen, X., Xu, D., et al, Forecasting acute hypotensive episodes in intensive care patients based on a peripheral arterial blood pressure waveform. In *Comput Cardiol*, 2009, 36.

[20] Saeed, M., Mark, R.G., A Novel Method for the Efficient Retrieval of Similar Multiparameter Physiologic Time Series Using Wavelet-base Symbolic Representations. In *AMIA Annu Symp Proc. 2006*, 2006, 679–683.

[21] Parati, G., Saul, J.P., Rienzo M.D., Mancia G., Spectral Analysis of Blood Pressure and Heart Rate Variability in Evaluating Cardiovascular Regulation. In *Hypertension*, 1995, 25, 1276-1286.

[22] Cao, H., Norris, P., et al. Mortality and Non-Invasive Measurement of Myocardial Oxygen Consumption Variability: A Study of 935 Trauma Patients. In *The 29th Annual Conference on Shock 2007*, Baltimore, MD, Jun 9-12, 2007.

[23]   Rosner, M.J., Rosner, S.D., Johnson, A.H.. Cerebral perfusion pressure: management protocol and clinical results. *J Neurosurg,* 1995, 83, 949-962

[24]   Guidelines for the management of severe head injury. Brain Trauma Foundation, American Association of Neurological Surgeons, Joint Section on Neurotrauma and Critical Care. *J Neurotrauma,* 1996, 13, 641-734.

[25]   Stoker, M.R., Principles of pressure transducers, resonance, damping and frequency response, *Anaesthesia and intensive care medicine*, 2004, 5(11), 371-5.

[26]   Cao, H., Norris, P., et al, A Simple Real-time Non-physiological Artifact Filter for Invasive Arterial Blood Pressure Monitoring: a Study of 1852 Trauma ICU Patients. In *Conf Proc IEEE Eng Med Biol Soc. 2006*, 1, 1417-20