# ForecastICU: A Prognostic Decision Support System for Timely Prediction of Intensive Care Unit Admission

Jinsung Yoon, Ahmed M. Alaa, *Member, IEEE,* Scott Hu, *MD* and Mihaela van der Schaar, *Fellow, IEEE*

**Abstract**

Intensive Care Unit (ICU) admission of hospitalized patients, for whom multiple physiological data streams are being monitored, is a vital and delay-critical clinical decision that has a direct impact on morbidity, mortality, and resource utilization. In order to improve the timeliness and accuracy of ICU admission decisions, which mainly rely on physicians, we developed *ForecastICU*: a prognostic decision support system that monitors hospitalized patients and prompts alarms for ICU admission. ForecastICU is first trained in an offline stage by constructing a Bayesian *belief system* that corresponds to its belief about how trajectories of physiological data streams of the patient map to a clinical status. After that, ForecastICU monitors a new patient in real-time by observing her physiological data stream, updating its *belief* about her status over time, and prompting an alarm whenever its *belief process* hits a predefined threshold (confidence). Using a real-world dataset obtained from a large academic medical center, we show that ForecastICU can predict ICU admissions 9 hours before a physician's decision (for a sensitivity of 40% and a precision of 50%). Moreover, ForecastICU performs consistently better than other benchmark algorithms in terms of sensitivity, precision and timeliness for all settings of the system parameters: it can predict ICU admissions 3 hours earlier, and offers a 7.8% gain for sensitivity, 5.1% gain for precision, and a 8.15% gain for the area under curve compared to the best benchmark.

**Index Terms**

Bayesian learning, Intensive care unit, Prognosis, Prediction, Time-series classification.

J. Yoon, A. M. Alaa, and M. van der Schaar are with the Department of Electrical Engineering, University of California Los Angeles, UCLA, Los Angeles, CA, 90024, USA (email: jsyoon0823@ucla.edu, ahmedmalaa@ucla.edu, mihaela@ee.ucla.edu).

S. Hu is with the David Geffen School of Medicine, University of California Los Angeles, UCLA, Los Angeles, CA, 90024, USA (email: scotthu@mednet.ucla.edu).

# ForecastICU: A Prognostic Decision Support System for Timely Prediction of Intensive Care Unit Admission

## I. Introduction

Intensive Care Unit (ICU) admission for hospitalized patients is a vital and delay-critical decision. The timing of transfer to the ICU is an important determinant of a patient's outcomes. Various medical studies have demonstrated that delayed identification of clinical deterioration, leading to delayed admission to the ICU and delayed therapeutic intervention, results in increased morbidity and mortality [1] [2]. With increasing physician and nursing workloads and more hand-offs of care, prompt recognition of a deteriorating patient has become increasingly difficult. This illustrates the urgency to develop automated prognostic decision support systems that alert the medical staff of impending clinical deterioration, enabling clinicians to intervene at an earlier time, thereby preventing an arrest or reducing the need for ICU transfer.

An automated system for ICU admission prediction is envisioned to operate in the following manner. The system will be fed with high-dimensional physiological data streams that belong to a monitored, hospitalized patient. It will try to infer whether the monitored patient is clinically deteriorating or not in a timely manner, i.e. earlier than the time at which a physician would normally decide to impend an ICU transfer for that patient. The system can take advantage of the available electronic health record (EHR) data in order to learn the trends in the physiological data streams associated with patients who previously got discharged or admitted to the ICU [3].

Designing a system that carries out the steps described above is associated with many practical and technical challenges. First, while data streams for previously hospitalized patients is recorded in the EHR, the *clinical status* of such patients upon their hospitalization *differs* from one patient to another, thus learning from the labeled patient examples is not straightforward as the examples themselves entail some ambiguity. Second, *not all* the data streams are *relevant* to the ICU admission decision, and not all of the different streams are sampled with the same rate, which implies that some values in some of the streams will be missing. Finally, the patients are monitored only for a finite amount of time, and the time-series observed by the system would eventually stop, thus the system should issue the prediction with a reasonable amount of time ahead of an *unknown deadline*, i.e. the actual time when physicians decide to admit the patient to ICU or discharge her.

In this paper, we develop ForecastICU, a prognostic decision support system that carries out timely predictions of ICU admissions for hospitalized patients. ForecastICU adopts a Bayesian approach for issuing predictions; by applying density estimation using the data streams of previously hospitalized patients, ForecastICU constructs a *belief system* that corresponds to its belief about the patient's clinical status as a function of time. ForecastICU approaches a new patient by observing her physiological data stream, updates its *belief* about the patient over time

as it observes more samples from her physiological stream, and prompts an ICU admission alarm whenever the system's *belief process* hits a predefined threshold that quantifies the system's confidence in the issued alarms.

Our approach hinges on the idea that predicting ICU admissions from temporal physiological streams can be viewed as an optimal stopping problem; or in other words, it is equivalent to learning an *unknown stopping rule* of a stochastic process, i.e. learning how physicians make ICU admission decisions from the recorded temporal data streams (realizations of a stochastic process) of previously hospitalized patients, and hence promptly issuing ICU admission predictions (alarms) before the stopping times of these processes (i.e. before the ICU admission decision that physicians would actually make without ForecastICU). Technically, ForecastICU approaches the ICU prognosis problem as an optimal stopping problem with uncertainty in the initial clinical status and the distribution of the physiological data streams.

We have applied ForecastICU to a real-world dataset obtained from a large academic medical center. Experiments show that, for a sensitivity of 40% and a precision of 50%, ForecastICU can predict ICU admissions as early as 9 hours (on average) before the actual physician's decision. Moreover, we show that ForecastICU performs consistently better than other benchmark algorithms (including random forest, LASSO, logistic regression, and SVMs) for all ranges of the system's parameters in terms of sensitivity, precision and timeliness: it can predict ICU admissions 3 hours earlier, and offers a 7.8% gain for sensitivity and 5.1% gain for precision with a 8.15% gain for the area under curve (AUC) compared to the best benchmark algorithm. Such gains can map to significant reductions in ICU mortality rates and better resource utilization in hospitals.

## II. Related works

We classify the related works into three categories: previously established methodologies for ICU prognosis in both the data mining and medical literature, general algorithms developed for time-series classification, and previous works that utilized the Bayesian learning framework.

### A. ICU prognosis

Methods for supporting prognostic clinical decisions have been investigated both in the medical literature and in the data mining literature. Several clinical studies have investigated the effectiveness of the usage of early warning scores (EWS) for the detection of patient deterioration [4]. Such methods identifies hospitalized patients who reach a certain "trigger" threshold [3] [5], and consequently alarms the clinicians to impend an ICU transfer for those patients. However, since EWS-based methods respond to triggering events that may not signal a truly deteriorating patient, they suffer from high rates of false alarms (70-95%) [6], which results in alarm fatigue and inappropriate resource utilization. Recent systematic reviews have demonstrated that EWS-based alarms only marginally improve outcomes while substantially increasing physician and nursing workloads [7].

Prognostic decision support has been investigated in the data mining literature as well. In [8], [9], and [10], simple regression models where developed to carry out risk assessment for developing diseases like breast and prostate cancer based on the patients' features. However, such predictive models deal with scenarios where predictions span years rather than hours, i.e. the delay-sensitivity of such models is much coarser and less critical. Prognostic

decision support for delay-critical settings has been addressed for the context of ICU environments in [11]. [12] and [13]. However, the objective in these models was to predict the trajectory of physiological data streams using previously hospitalized patients' synchronized data streams. Unlike ForecastICU, the models therein have developed simple auto-regressive models to extrapolate incoming data streams, but have not addressed the problem of timely classification of incoming patients whose monitored data streams entail an unknown stopping time, and are not synchronized with the training data.

### *B. Time-series classification*

A wealth of literature on time-series classification exists. Most of the time-series classification algorithms rely on clustering the time-series in the training data using a distance metric, and then associating incoming series with one of the clusters. For instance, a 1-nearest neighbor approach was used in [14] and [15], a semi-supervised approach employing Euclidean distance was used in [16], clustering using dynamic time warping (DTW) as a distance metric was adopted in [17], and clustering using a Kalman-filter-based approach was introduced in [18]. Methods based on recurrent neural networks [19], ensemble learning [20], joint classification-regression [21][22], or online prediction with missing data samples [23] where also investigated. Moreover, various methods for predicting trends of online content where developed, i.e. in [24] the problem of forecasting the popularity of videos was addressed, whereas predicting trends in Twitter was investigated in [25] and [26].

Our work departs from this literature in the following aspects. First, unlike the conventional time-series classification problem, the physiological data stream stops at some point of time (when the patient is discharged or admitted to ICU), so the goal is not just to issue an early prediction, but to prompt an alarm that precedes the stopping time of the stream with a long enough time interval. Furthermore, our training data comprises data streams for patients who were hospitalized with different clinical statuses, thus ForecastICU needs to learn how to "align" the training data in order to construct its belief system. Finally, since ForecastICU prompts alarms to physicians and nurses, it needs to guarantee a desirable level of "confidence" in its predictions, which was not considered in any of these works.

### *C. The Bayesian Learning Framework*

ForecastICU adopts a Bayesian learning framework; by that we mean: it constructs a belief system about how the status of an ICU patient changes over time using the training data, and applies Bayesian updates for its posterior belief about the status of monitored patients. The Bayesian learning framework has natural connections with human cognition [27] [28], and has been extensively studied in the economics literature [29] [30]. Bayesian learning has been studied thoroughly in that literature assuming that the density function based on which posterior beliefs are updated is perfectly known, and the goal there is to see whether or not an unknown state-of-the-world (hypothesis) can be learned over time [31]. The same assumptions were adopted in the optimal stopping problems considered in the finance literature [32]. However, for the problem of ICU admission prediction, the density function of the patients' physiological data streams is unknown, their status at hospitalization is ambiguous and the goal is to learn in a *timely* manner the state of that patient. This imposes a set of very different challenges on our learning problem

that is not encountered in the context of theoretical economics, thus new analyses and learning algorithms that can deal with these challenges are needed.

## III. Theoretical foundations for ForecastICU

In this section, we provide a formal model for an abstract *forecaster* that aims at learning to predict ICU admissions. The theoretical analysis of such a forecaster will serve as the foundation for the practical forecaster implementation, the ForecastICU algorithm, which we will present in the next section.

### A. Mathematical model

Given a *probability space* $(\Omega, \mathcal{F}, \mathbb{P})$, and a measurable space $(\mathcal{S}, \Psi)$, a stochastic process $\mathbf{X}_\mathcal{T}$ comprises a collection of $\mathcal{S}$-valued random variables on $\Omega$ indexed by a totally ordered set $\mathcal{T}$. That is, $\mathbf{X}_\mathcal{T} = \{X_t\}_{t\in\mathcal{T}}$, where every $X_t$ is an $\mathcal{S}$-valued random variable on $\Omega$. Given the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a *filtration* $\{\mathcal{F}_t, t \in \mathcal{T}\}$ is a (weakly) increasing collection of $\sigma$-algebras on $\Omega$. The filtration $\mathcal{F}_t$ is always bounded above by $\mathcal{F}$, i.e. $\mathcal{F}_t \subseteq \mathcal{F}$. The process $\mathbf{X}_\mathcal{T}$ is a *martingale* with respect to the filtration $\mathcal{F}_t$ if $\mathbb{E}[X_{t+1}|\mathcal{F}_t] = X_t$. The *stopping time* $\tau_s$ of the process $\mathbf{X}_\mathcal{T}$ is a random variable $\tau_s : \Omega \to \mathcal{I}$, where $\mathcal{I}$ is an ordered index set, e.g. $\mathcal{I} = [0, \infty)$. A stopping time $\tau_s$ satisfies that $\{\omega \in \Omega : \tau_s(\omega) \leq t\} \in \mathcal{F}_t, \forall t \in \mathcal{I}$. The stopping time of a process is decided by some arbitrary *stopping rule*. The *hitting time* $\tau_h$ of a process $\mathbf{X}_\mathcal{T}$ is the first time it hits a certain value, i.e. $\tau_h(\eta) = \inf\{t \in \mathbb{R} | X_t \geq \eta\}$.

We consider every patient's physiological data stream as a stochastic process with respect to the space $(\Omega, \mathcal{F}, \mathbb{P})$. For instance, the blood pressure measurements stream can be viewed as a stochastic process $\mathbf{X}_\mathcal{T}$ that is observed starting from the time the patient was hospitalized. The stopping time $\tau_s$ of such a process is the time at which the physician takes a decision regarding the patient: the decision can be either discharging the patient or admitting her to the ICU, where in both cases the patient is not monitored further. The stopping rule that determines such a stopping time is simply the physician's criteria for ICU admission or discharging of hospitalized patients, which depends on how physicians interpret the physiological data stream.

Patients belong to two categories: stable patients who should be discharged, and clinically deteriorating patients who should be admitted to the ICU. We assume that the null hypothesis $\mathcal{H}_0$ is the hypothesis that the patient is stable, whereas the alternative hypothesis $\mathcal{H}_1$ is the hypothesis that the patient is clinically deteriorating. Depending on whether the true hypothesis is $\mathcal{H}_0$ or $\mathcal{H}_1$, a physiological data stream $\mathbf{X}_\mathcal{T}$ will have a different joint distribution for its data samples with respect to the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We denote the families of finite-dimensional distributions of the physiological streams under the null and alternative hypotheses as $\mathbb{P}_0$ and $\mathbb{P}_1$ respectively, i.e. $\{X_\tau\}_{\tau=0}^t | \mathcal{H}_m \sim \mathbb{P}_m^t, m \in \{0, 1\}$, and $\mathbb{P}_m$ is the family of distributions $\mathbb{P}_m^t$ for all admissible values of $t$ (e.g. maximum time a patient can stay hospitalized).

### B. The Forecaster

*1) Formal definition:* We formally define the *forecaster* as a *belief system* that carries out the mapping $B_t : (\mathcal{F}_t, \mathbb{Q}_0, \mathbb{Q}_1) \to [0, 1]$, i.e. a map from a filtration to a *belief* about the monitored patient being clinically deteriorating,

where $\mathbb{Q}_0$ and $\mathbb{Q}_1$ are estimates of the families of finite-dimensional distributions $\mathbb{P}_0$ and $\mathbb{P}_1$. We say that the belief system is *truthful* if $\mathbb{Q}_m = \mathbb{P}_m, m \in \{0,1\}$, and we say that it is *non-truthful* otherwise, i.e. the estimated densities $\mathbb{Q}_m$ of a non-truthful belief system have a non-zero distance from $\mathbb{P}_m$ with respect to any probability metric. The function $B_t$ (which we will also write as $B_t\left(\mathcal{H}_1 \,|\mathcal{F}_t\,\right)$) is denoted as the *belief function*, whereas the sequence $\{B_t\}_{t\in\mathcal{T}}$ is called the *belief process*. Informally, the forecaster is endowed with some information about the physiological stream generative process, encoded in the distributions $\mathbb{Q}_0$ and $\mathbb{Q}_1$, and it accumulates information over time as it observes the monitored patient's physiological streams, and builds a belief about her clinical status. Intuitively, the accuracy and timeliness of the forecaster depend on the quality of the estimates $\mathbb{Q}_0$ and $\mathbb{Q}_1$, and the way the mapping $B_t$ is implemented given the filtration $\mathcal{F}_t$.

*2) Structure of the forecaster:* The implementation of the forecaster's belief system demands two basic modules: a *density estimation algorithm* $\mathcal{A}^D$, which finds "good" estimates $\mathbb{Q}_0$ and $\mathbb{Q}_1$ in an offline manner, and a *belief function* $B_t$, which updates the forecaster posterior belief in real-time in response to the information extracted from the monitored data streams. Formally, given that the forecaster has access to a dataset $\mathbf{X}^{ref}$ comprising recorded data streams of $N$ reference patients in the EHR who are labeled as being admitted to ICU or discharged, the density estimation algorithms is a mapping $\mathcal{A}^D : \mathbf{X}^{ref} \rightarrow (\mathbb{Q}_0, \mathbb{Q}_1)$, and the belief function is a real-time mapping $B_t : (\mathcal{F}_t, \mathcal{A}^D\left(\mathbf{X}^{ref}\right)) \rightarrow [0,1]$.

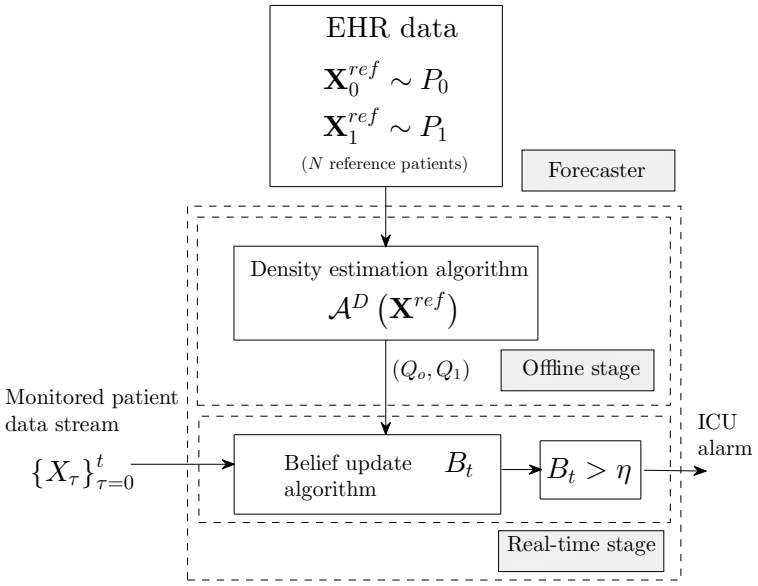


Fig. 1. Schematic for a forecaster that learns to issue ICU admission alarms from the EHR data.

*3) Alarm strategy and performance:* The problem of (timely) predicting ICU admissions can be thought of as being equivalent to an optimal stopping problem, or a problem of *learning an unknown stopping rule* of a stochastic process. Thus, not only does the forecaster face uncertainty in the true hypothesis, but also it is uncertain about when will the process stop; the forecaster needs to figure out the true hypothesis before the process stops with a reasonable amount of time.

ForecastICU adopts a "threshold type" *alarm strategy*: it prompts an alarm for an ICU admission whenever its belief process first crosses a predefined threshold $\eta$. Thus, the optimal stopping time problem boils down to finding the optimal threshold that the belief process should hit before an ICU alarm is issued. Selection of the threshold $\eta$ is aimed to maximize a *clinical value function* that comprises a set of accuracy and timeliness measures.

Three measures of performance are clinically relevant to the ICU prognostic setting: timeliness, sensitivity and precision [33] [5] [7]. Let $\tau_s$ be the stopping time of the patient's physiological data stream, and let $\tau_h(\eta)$ be the hitting time of the belief process $\{B_t\}_{t\in\mathbb{N}}$ given a threshold $\eta$. The sensitivity of the forecaster, which is also known as the *true positive rate* (TPR), is given by

$$TPR = \frac{\mathbb{P}\left(\tau_h(\eta) < \tau_s \,|\mathcal{H}_1\right)}{\mathbb{P}\left(\tau_h(\eta) < \tau_s \,|\mathcal{H}_1\right) + \mathbb{P}\left(\tau_h(\eta) > \tau_s \,|\mathcal{H}_1\right)}, \tag{1}$$

whereas the precision, which is also known as the *positive predictive value* (PPV), is given by

$$PPV = \frac{\mathbb{P}\left(\tau_h(\eta) < \tau_s \,|\mathcal{H}_1\right)}{\mathbb{P}\left(\tau_h(\eta) < \tau_s \,|\mathcal{H}_1\right) + \mathbb{P}\left(\tau_h(\eta) < \tau_s \,|\mathcal{H}_0\right)}. \tag{2}$$

Finally, the timeliness of the forecaster $T_p(\eta)$ is the average time interval between the hitting time of the belief process and the stopping time of the physiological data stream for clinically deteriorating patients, which is formally given by

$$T_p(\eta) = \mathbb{E}\left[\tau_s - \tau_h(\eta) \,|\tau_h(\eta) < \tau_s, \mathcal{H}_1\right]. \tag{3}$$

The selection of the threshold value $\eta$ should balance the trade-off between accuracy (in terms of TPR and PPV) and timeliness; intuitively, one expects that low threshold values would lead to more timely but less accurate decisions, and vice versa. Formally, we define a general clinical reward function $g\left(\{X_\tau\}_{\tau=0}^{\tau_h(\eta)}\right)$ that quantifies the overall performance in terms of PPV, TPR and $T_p$ as a function of the alarm strategy $\eta$. The ICU prognostic optimal stopping problem is equivalent to finding an optimal alarm strategy (a threshold $\eta^*$) that maximizes a *clinical value function* $V(\eta)$ as follows

$$V^* = \sup_{\eta\in[0,1]} \mathbb{E}_{\mathbb{P}}\left[g\left(\{X_\tau\}_{\tau=0}^{\tau_h(\eta)}\right)\mathbf{1}_{\{\tau_h(\eta)<\tau_s\}}\right], \tag{4}$$

where the optimal clinical value function is $V^* = V(\eta^*)$. Problem (4) is challenging for that $\mathbb{P}$ is unknown to the forecaster, and $\tau_s$ is random. The optimal solution to (4) would balance the value of information in the physiological stream $\{X_\tau\}_{\tau=0}^{\tau_h(\eta)}$ and the TPR, PPV and $T_p$ which are reflected in the reward function $g$. Fig. 1 illustrates the structure of the forecaster: its offline and real-time components, and its alarm strategy.

### C. Bayesian Learning

The forecaster's alarm strategy is a threshold strategy on its belief process. The forecaster builds the belief process given observed information using a Bayesian learning approach: it updates its posterior belief in response to new observations of the physiological data streams as follows

$$B_t\left(\mathcal{H}_1 \,|\mathcal{F}_t\right) = \frac{\mathbb{Q}\left(\{X_\tau\}_{\tau=0}^{t} \,|\mathcal{H}_1\right)\mathbb{Q}\left(\mathcal{H}_1\right)}{\sum_{i\in\{0,1\}}\mathbb{Q}\left(\{X_\tau\}_{\tau=0}^{t} \,|\mathcal{H}_i\right)\mathbb{Q}\left(\mathcal{H}_i\right)},$$

where $\{\mathbb{Q}(\mathcal{H}_m)\}_{m\in\{0,1\}}$ are the forecaster's estimates of the discrete priors. The forecaster cannot compute the optimal threshold in (4) since the belief system is generally non-truthful. Thus, the accuracy and timeliness of

the forecaster (the value function $V(\eta)$) will be affected by the truthfulness of its belief system. In the following Theorem, we link the truthfulness of the forecaster's belief system to the timeliness and accuracy of the alarm strategy in terms of the sample complexity. The proof of this Theorem is given in the Appendix.

*Theorem 1: (Probably approximately correct and timely ICU alarm strategies)* For every $(\epsilon, \delta) \in [0,1]^2$, there exists a polynomial function $N^*(\epsilon, \delta) = poly(\frac{1}{\delta}, \frac{1}{\epsilon})$ and a density estimation algorithm $\mathcal{A}^D$, such that for every dataset with $N > N^*(\epsilon, \delta)$ reference patients, we have that $\mathbb{P}\left(|V(\eta) - V^*| < \epsilon\right) \geq 1 - \delta$, where $V^*$ is optimal clinical value function of a truthful belief system that has access to the optimal threshold $\eta^*$.

Theorem 1 makes a link between our timely (binary) classification problem and the classical probably approximately correct (PAC) framework for binary classification. The two problems are profoundly different since the ICU prognosis problem entails timeliness and random stopping of data observation; however, a supervised learning guarantee can still be realized. The Theorem says that one can compute a sample complexity (the number of reference patients) that achieves an $\epsilon$-optimal clinical value function (compared to an "oracle" belief system that has access to the true data streams' distributions), with an arbitrary level of confidence $1 - \delta$. That is, the ICU prognosis problem using the proposed forecaster structure is *learnable* in the "probably approximately correct and timely" sense. This motivates the construction of a practical forecaster algorithm in the next section.

## IV. The ForecastICU Algorithm

Practical implementation of the forecaster is confronted with several challenges. First, not all the monitored physiological streams are relevant to the ICU prognostic decision. Second, the data streams are not sampled with the same rate. Finally, a crucial aspect of the temporal physiological data is that the data streams of the reference patients are neither synchronized with each other, nor they are synchronized with the incoming patients since each patient is hospitalized in a different clinical status.

As shown in Fig. 1, designing a forecaster entails designing a density estimation algorithm $\mathcal{A}^D(\mathbf{X}^{ref})$ and a belief updating procedure. In the following two subsections, we propose a design for both modules as the two building blocks of ForecastICU.

### A. The offline density estimation algorithm $\mathcal{A}^D(\mathbf{X}^{ref})$

The algorithm $\mathcal{A}^D(\mathbf{X}^{ref})$ takes as an input a dataset $\mathbf{X}^{ref}$ with $N$ reference patient entries, associated with each patient a set of $L$ recorded physiological data streams, and retrieves the estimated finite-dimensional distributions $(\mathbb{Q}_0, \mathbb{Q}_1)$. We denote the $i^{th}$ reference patient by $\mathbf{X}^{ref}_{(i)}$, where $\mathbf{X}^{ref}_{(i)}$ is an $L \times K_i$ matrix, with $K_i$ being the length of the longest data stream associated with reference patient $i$. We denote the overall hospitalization period of patient $i$ as $T^H_{(i)}$. The algorithm implements the following four steps:

**a- Non-causal alignment of reference patients' data streams :** Since the reference patients are hospitalized with different statuses, ForecastICU needs to align their data streams prior to applying density estimation. The stopping time of the data streams usually follows a rational, consistent stopping rule decided by the physicians, and such a rule serves as a marker for the clinical status of the patient at the ICU admission time or discharge time. Therefore, ForecastICU aligns the patients' data streams in a non-causal fashion: it views the stopping times of all streams

in the dataset $\mathbf{X}^{ref}$ as the reference time ($t = 0$) of the corresponding stochastic processes. We use the notation $\mathbf{X}^{ref}_{(i)}(m, n)$ to refer to the data sample of the $m^{th}$ data stream at $n$ time steps ahead of the (reference) stopping time.

**b- Interpolating under-sampled data streams :** The different data streams for the same patient can be sampled with different sampling rates. Thus, the algorithm $\mathcal{A}^D(\mathbf{X}^{ref})$ would encounter missing values at some points in time when carrying out the joint density estimation across the features. To reconstruct the missing samples in under-sampled data streams, we use an *interpolation function* $h\left(\left\{\mathbf{X}^{ref}_{(i)}(m,n)\right\}_{n=0}^{K-1}, T_s\right) : \mathbb{R}_{K\times 1} \times \mathbb{R} \rightarrow \mathbb{R}_{\lfloor \frac{KT_s}{T^H_{(i)}} \rfloor \times 1}$, i.e. a function that interpolates samples of a data stream and retrieves a stream with a sampling period of $T_s$ instead of $\frac{T^H_{(i)}}{K}$. ForecastICU uses *cubic spline* interpolation function that we denote by $h_{spline}$. We denoted the interpolated dataset as $\tilde{\mathbf{X}}^{ref} = h_{spline}\left(\mathbf{X}^{ref}\right)$, where all the data streams in $\tilde{\mathbf{X}}^{ref}$ are up-sampled with the sampling rate of the most frequently sampled data stream.

**c- Selecting relevant data streams :** ForecastICU applies correlation feature selection (CFS) algorithm on the interpolated dataset $\tilde{\mathbf{X}}^{ref}$ (with minimum-redundancy-maximum relevance (mRMR) [34]) to discover the relevant temporal data streams for forecasting ICU patients. We denote the data stream selection operation as $\tilde{\mathbf{Y}}^{ref} = CFS\left(\tilde{\mathbf{X}}^{ref}\right)$, where $\tilde{\mathbf{Y}}^{ref}$ is a dataset with $N$ reference patients, associated with each a set of $R \leq L$ data streams, where $R$ is the number of discovered relevant data streams.

**d- Parametric density estimation :** We estimate the finite-dimensional distributions $(\mathbb{Q}_0, \mathbb{Q}_1)$ based on the processed dataset $\tilde{\mathbf{Y}}^{ref}$ as follows. Since the dataset is labeled, we separate the reference patients into clinically deteriorating patients with processed data streams $\tilde{\mathbf{Y}}^{ref}_1$, and clinically stable patients with data streams $\tilde{\mathbf{Y}}^{ref}_0$. Let $N_0$ and $N_1$ be the number of entries in $\tilde{\mathbf{Y}}^{ref}_0$ and $\tilde{\mathbf{Y}}^{ref}_1$ respectively, and $K_0$ and $K_1$ be the average lengths of the data streams in $\tilde{\mathbf{Y}}^{ref}_0$ and $\tilde{\mathbf{Y}}^{ref}_1$ respectively. Let $(\mathbb{Q}^t_0, \mathbb{Q}^t_1)$ be the joint density functions of all the data samples for the $R$ data streams selected by CFS between the (average) hospitalization time $K_0$ (or $K_1$) and the a time instance that is $t$ steps away from the hospitalization time. We approximate the finite-dimensional distributions $(\mathbb{Q}^t_0, \mathbb{Q}^t_1)$ as Multi-variate Gaussian distributions, and fit their mean and covariance parameters, i.e. $\mathbb{Q}^t_m \sim \mathcal{N}\left(\hat{\mu}^t_m, \hat{\mathbf{\Sigma}}^t_m\right), m \in \{0, 1\}$ where

$$\left[\hat{\mu}^t_m(j)\right]_{j=1}^{R} = \frac{1}{N_m}\sum_{i=1}^{N_m} \tilde{\mathbf{Y}}^{ref}_{(i),m}(j, t), \tag{5}$$

$$\left[\hat{\mathbf{\Sigma}}^t_m\right]_{k,l} = \frac{1}{N_m - 1}\sum_{i=1}^{N_m} \bar{\mathbf{Y}}^{ref}_{(i),m}(k, t)\bar{\mathbf{Y}}^{ref}_{(i),m}(l, t), \tag{6}$$

where $k, l = 1, \ldots, R$, and $\bar{\mathbf{Y}}^{ref}_{(i),m}(k, t) = \tilde{\mathbf{Y}}^{ref}_{(i),m}(k, t) - \frac{1}{N_m}\sum_{w=1}^{N_m} \tilde{\mathbf{Y}}^{ref}_{(w),m}(k, t)$. Thus, the output of $\mathcal{A}^D(\mathbf{X}^{ref})$ is a family of estimated finite-dimensional distributions $(\mathbb{Q}_0, \mathbb{Q}_1)$, i.e. a set of Multi-variate Gaussian densities for different values of the elapsed hospitalization period. The density estimation algorithm passes the density estimates to the belief function which runs in real-time and applies Bayesian learning using $(\mathbb{Q}_0, \mathbb{Q}_1)$ as priors with which it updates its posterior beliefs.

### B. The real-time belief updating algorithm

The belief function of ForecastICU is updated at time $t$ given a the data stream $\{X_\tau\}_{\tau=t_o}^{t}$ (or generally, the information $\mathcal{F}_t$ available at time $t$) of the incoming patient as follows $B_t\left(\mathcal{H}_1 \left| \mathcal{F}_t \right.\right) = \mathbb{Q}\left(\mathcal{H}_1 \left| \{X_\tau\}_{\tau=t_o}^{t} \right.\right)$. However, since ForecastICU faces uncertainty about the clinical status, the reference time $t_o$ of the stream $\{X_\tau\}_{\tau=t_o}^{t}$ with respect to the estimated densities $(\mathbb{Q}_0, \mathbb{Q}_1)$ is unknown. Therefore, ForecastICU estimates the belief function as

$$\frac{N_1 \mathbb{Q}\left(\{X_\tau\}_{\tau=t_o^1}^{t} | \mathcal{H}_1, T_1^*(t)\right)}{N_0 \mathbb{Q}\left(\{X_\tau\}_{\tau=t_o^0}^{t} | \mathcal{H}_0, T_0^*(t)\right) + N_1 \mathbb{Q}\left(\{X_\tau\}_{\tau=t_o^1}^{t} | \mathcal{H}_1, T_1^*(t)\right)},$$

where $t_o^m = t - T_m^*(t) + 1$, and $T_m^*(t)$ is the "most likely" estimate for the time remaining until the stopping time of the process, which is simply given by

$$T_m^*(t) = \arg\max_{\tau} \mathbb{Q}\left(\{X_k\}_{k=\tau-t+1}^{\tau} | \mathcal{H}_m\right). \tag{7}$$

In order to rule out drastic fluctuations and spikes in the belief function, which may result from a belief system constructed from a small data set, we apply a smoothing phase for the belief process via a simple moving average filter of length $W$. The smoothed belief function is given by $\tilde{B}_t = \frac{1}{W}\sum_{\tau=t-W}^{t} B_\tau$. ForecastICU prompts an ICU alarm whenever $\tilde{B}_t$ exceeds a threshold $\eta$. In the following subsection, we show how the threshold $\eta$ is set.

### C. The alarm strategy

ForecastICU follows a threshold-type alarm strategy: it prompts an ICU alarm whenever the smoothed belief process $\tilde{B}_t$ hits a threshold $\eta$. The fundamental trade-off that ForecastICU balances is the one between the timeliness of a prediction and its accuracy. We control such a trade-off via two parameters: the threshold $\eta$ and the size of the smoothing filter $W$. In this case, the alarm strategy is defined by a richer set of parameters $(\eta, W)$ rather than being solely determined by $\eta$. The threshold $\eta$ controls to the confidence in the issued alarms, and $W$ controls the stability of the belief function's fluctuations, and hence the system's sensitivity. In order to select desirable values for $\eta$ and $W$, we define a clinical value function $V(\eta, W)$ as follows

$$V(\eta, W) = \alpha\, TPR(\eta, W) + (1-\alpha) T_p(\eta, W) \tag{8}$$

where $\alpha \in [0, 1]$ is a parameter that balances the preferences over timeliness and sensitivity. By setting a constraint $\gamma$ on the system's precision (which can be determined by physicians), ForecastICU computes the alarm strategy by solving the following optimization problem:

$$\begin{aligned} \max_{\eta, W} \quad & V(\eta, W) \\ \text{subject to} \quad & PPV(\eta, W) \geq \gamma. \end{aligned} \tag{9}$$

## V. Experiments

### A. Data Description

ForecastICU was applied to a population of patients admitted to a large academic hospital. The population considered is fairly homogeneous: most patients were diagnosed with leukemia, lymphoma, multiple myeloma

and other hematologic malignancies. The majority of the patients were receiving chemotherapy, allogeneic stem cell transplantation or autologous stem cell transplantation. We have chosen this particular population of patients for our experiments because these patients were receiving treatments (according to their diagnosis) that cause severe immunosuppression during their hospitalization, placing them at an extreme risk of developing a clinical deterioration, which requires ICU admission. This is extremely important as delays in ICU admission in these patients are associated with increased mortality and morbidity. [1] [35] [2] [33].

The patients' clinical features comprise 18 temporal physiological data streams which are described in the Appendix. Each patient's data stream is associated with a binary label: either discharge (DIS) or ICU Admission (ICU). The total number of patients is 1065, and the number of patients admitted to ICU is 101 (9.48%). The hospitalization period ranges from 4 to 2062 hours (85 days and 22 hours).

### *B. Experiments Setup*

We compare the performance of ForecastICU in terms of timeliness, PPV and TPR with respect to four state-of-the-art machine learning techniques: logistic regression (LR), support vector machines with radial based kernel (SVMs), regularized logistic regression with Lasso (LASSO) and random forest. Performance measures were computed via 10-fold stratified cross validation: we run 10 independent cross validations and report the average as the final performance. Clinicians responsible of ICU admissions in the medical center from which we extracted the data indicated that ICU alarms would be most helpful if they are at least 4 hours earlier than the standard time an ICU admission decision would be taken by the clinician, in order to provide sufficient time to safely enact the transfer and to potentially correct the cause of the underlying clinical deterioration.

### *C. Experiment Results*

*1) Performance Comparison with Benchmarks:* ForecastICU consistently outperforms the other benchmark algorithms with respect to the PPV for every given value of the TPR as shown in Fig. 2 and Table I. For instance, for a TPR of 50%, ForecastICU achieves a PPV of 54.7%, which is 5.1% better than the best benchmark algorithm (random forest). Moreover, given for a PPV of 40%, Forecast ICU achieves a TPR of 68.1%, which is 7.8% better than the best benchmark algorithm, with 8.15% gain in terms of the area under curve (AUC) as well. Note that while for different ranges of TPR and PPV, the best benchmark algorithm changes, ForecastICU is consistently outperforming that best benchmark.

Our algorithm offers a consistent improvement in terms of the rate of "false alarms" and "true alarms" as compared to the benchmarks; this is crucial in a practical ICU setting as studies suggest that the clinicans' response to an alarm is related to their confidence in the signal, which is directly related to how often the alarm correctly warns the clinician of impending danger [36] [37] [38]. For instance, at a TPR of 50%, ForecasetICU is the only algorithm among those being compared that could correctly alarm for an ICU admission with accuracy above 50%, whereas other benchmarks perform worse than what a physician could do with a simple coin flip. The $p$-value of the hypothesis test that compares ForecastICU and the best algorithm is less than $0.01$. Moreover, our algorithm can

display the belief threshold for the computed alarm strategy, which itself is a natural measure for the "likelihood" of the alarm being truthful, and is thus a measure of confidence in the issued alarms.

TABLE I

ACCURACY OF ICU ALARMS FOR FORECASTICU AND THE BENCHMARK ALGORITHMS.

| **Algorithms** | **TPR(%)** | **PPV(%)** |
|---|---|---|
| ForecastICU | 50.2± 1.45% | 54.7± 1.73% |
| Logistic Regression | 50.8± 1.77% | 39.7± 1.95% |
| Lasso Regularization | 50.7± 2.01% | 42.5± 1.98% |
| Random Forest | 51.1± 2.03% | 49.6± 1.55% |
| SVMs | 50.5± 2.12% | 29.8± 1.55% |

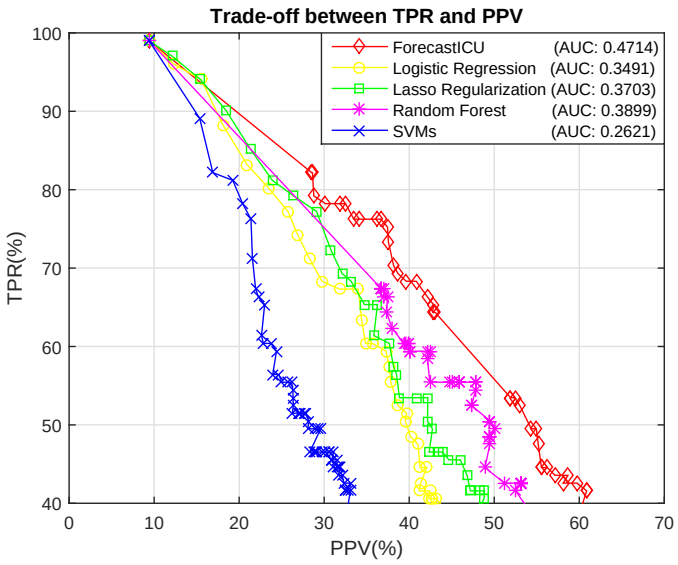


Fig. 2. Trade-off between TPR and PPV

*2) Trade-off between timeliness and accuracy:* Fig. 3 and Table II illustrate the trade-off between prediction time and its accuracy. While the performance of all algorithms naturally degrades as the ICU alarms are prompted at earlier times, we see that ForecastICU consistently outperforms all the other benchmark algorithms in terms of both TPR and PPV for all alarm times (every level of timeliness). For instance, the TPR of ForecastICU is 59.2% with a PPV of 30.3% for a 12-hour early prediction with respect to the actual physician-determined ICU admission event. This represents a gain of 3.4% with respect to the best benchmark algorithm, which in this case is the LASSO regularization. Fig. 3 also shows that ForecastICU can consistently predict the ICU admission earlier than the best benchmark algorithm, raising the alarm around 2 to 3 hours earlier. It can also predict ICU admissions 9 hours before a physician's decision for a PPV of 40% and a TPR of 50%.

TABLE II

ACCURACY OF ICU ADMISSION ALARMS BEFORE 12 HOURS FROM THE ACTUAL PHYSICIAN-DETERMINED ICU ADMISSION TIME

| Algorithms | TPR(%) | PPV(%) |
|---|---|---|
| ForecastICU | 59.2% | 30.3% |
| Logistic Regression | 55.7% | 30.7% |
| Lasso Regularization | 55.8% | 30.3% |
| Random Forest | 44.5% | 31.1% |
| SVMs | 32.3% | 29.9% |

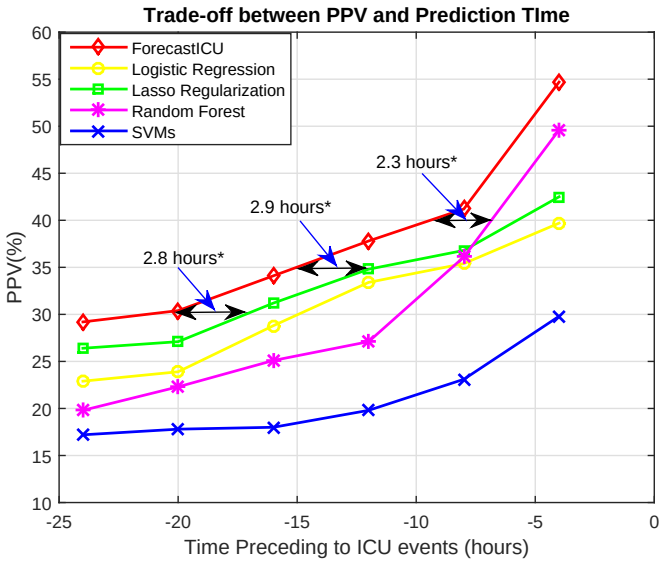


Fig. 3. Trade-off between PPV and the prediction time (TPR is fixed at 50%). (*: Earliness of ForecastICU alarms compared to the best benchmark algorithm)

### *D. Clinical significance of ForecastICU*

The clinical significance of ForecastICU is not only limited to the performance gains discussed earlier, but it also extends to its ability to handle a versatile clinical value function, which can assist clinicians in managing the ICU admission procedure. For instance, given a TPR determined by the clinician, ForecastICU is able to warn the clinician earlier and provide a more confident signal than other existing machine learning algorithms, thus providing the busy clinician with a safety net for patient care by giving them sufficient time to intervene at an earlier time in order to prevent clinical deterioration. Moreover, given that the value of PPV is related to the confidence that the clinician has in the alarm, the clinician's confidence with ForecastICU would be higher than other off-the-shelf benchmark algorithms regardless of the sensitivity (TPR) that the clinician decides to set. The PPV gains achieved by ForecastICU also imply a decrease in the number of patients that are falsely identified as needing ICU admission, which would reduce the medical reverse effects and unnecessary costs [39].

## VI. Conclusion

In this paper, we developed ForecastICU: prognostic decision support systems that monitors hospitalized patients and prompts alarms for ICU admission. We viewed the problem of ICU prognosis as an optimal stopping time problem, where a forecaster has to trade-off the value of information gained from monitoring a patient's physiological data streams with the timeliness of a potential ICU alarm. We have proposed ForecastICU as a practical algorithm that solves this problem using Bayesian learning.

ForecastICU was applied to a real-world dataset from a large academic medical center, and we have shown that it consistently outperforms benchmark algorithms in terms of timeliness, sensitivity and precision. In particular, ForecastICU can prompt ICU alarms that are around 2 to 3 hours earlier than the best benchmark, and can also issue ICU alarms 9 hours before a physician's ICU admission decision for a PPV of 40% and a TPR of 50%. Sensitivity and precision gains of 7.8% and 5.1% respectively with respect to the best benchmark algorithm were reported.

## References


[1] V. Liu, P. Kipnis, N. W. Rizk, and G. J. Escobar, "Adverse outcomes associated with delayed intensive care unit transfers in an integrated healthcare system," *Journal of hospital medicine*, vol. 7, no. 3, pp. 224–230, 2012.

[2] M. P. Young, V. J. Gooder, K. Bride, B. James, and E. S. Fisher, "Inpatient transfers to the intensive care unit," *Journal of general internal medicine*, vol. 18, no. 2, pp. 77–83, 2003.

[3] M. M. Churpek, T. C. Yuen, S. Y. Park, R. Gibbons, and D. P. Edelson, "Using electronic health record data to develop and validate a prediction model for adverse outcomes on the wards," *Critical care medicine*, vol. 42, no. 4, p. 841, 2014.

[4] S. R. Bruijns, H. R. Guly, O. Bouamra, F. Lecky, and W. A. Lee, "The value of traditional vital signs, shock index, and age-based markers in predicting trauma mortality," *Journal of Trauma and Acute Care Surgery*, vol. 74, no. 6, pp. 1432–1437, 2013.

[5] D. R. Prytherch, G. B. Smith, P. E. Schmidt, and P. I. Featherstone, "Views towards a national early warning score for detecting adult inpatient deterioration," *Resuscitation*, vol. 81, no. 8, pp. 932–937, 2010.

[6] C. L. Tsien and J. C. Fackler, "Poor prognosis for existing monitors in the intensive care unit," *Critical care medicine*, vol. 25, no. 4, pp. 614–619, 1997.

[7] N. Alam, E. Hobbelink, A. van Tienhoven, P. van de Ven, E. Jansma, and P. Nanayakkara, "The impact of the use of the early warning score (ews) on patient outcomes: a systematic review," *Resuscitation*, vol. 85, no. 5, pp. 587–594, 2014.

[8] E. W. Steyerberg, M. J. Eijkemans, F. E. Harrell, and J. D. F. Habbema, "Prognostic modeling with logistic regression analysis in search of a sensible strategy in small data sets," *Medical Decision Making*, vol. 21, no. 1, pp. 45–56, 2001.

[9] H. Seker, M. O. Odetayo, D. Petrovic, and R. N. Naguib, "A fuzzy logic based-method for prognostic decision making in breast and prostate cancers," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 7, no. 2, pp. 114–122, 2003.

[10] E. W. Steyerberg, M. J. Eijkemans, F. E. Harrell, and J. D. F. Habbema, "Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets," *Statistics in medicine*, vol. 19, no. 8, pp. 1059–1079, 2000.

[11] J. Sun, D. Sow, J. Hu, and S. Ebadollahi, "A system for mining temporal physiological data streams for advanced prognostic decision support," in *Data Mining (ICDM), 2010 IEEE 10th International Conference on*. IEEE, 2010, pp. 1061–1066.

[12] C. Neti, S. Ebadollahi, J. Hu, D. Gotz, J. Sun, and D. Sow, "Predicting patient's trajectory of physiological data using temporal trends in similar patients: A system for near-term prognostics." 2010.

[13] Y. Zhang, S. Fong, J. Fiaidhi, and S. Mohammed, "Real-time clinical decision support system with data stream mining," *BioMed Research International*, vol. 2012, 2012.

[14] Z. Xing, J. Pei, and S. Y. Philip, "Early classification on time series," *Knowledge and information systems*, vol. 31, no. 1, pp. 105–127, 2012.

[15] ——, "Early prediction on time series a nearest neighbor approach," in *IJCAI*. Citeseer, 2009, pp. 1297–1302.

[16] L. Wei and E. Keogh, "Semi-supervised time series classification," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2006, pp. 748–753.

[17] K. Buza, A. Nanopoulos, and L. Schmidt-Thieme, "Time-series classification based on individualised error prediction," in *Computational Science and Engineering (CSE), 2010 IEEE 13th International Conference on*. IEEE, 2010, pp. 48–54.

[18] L. Li and B. A. Prakash, "Time series clustering: Complex is simpler!" in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 2011, pp. 185–192.

[19] M. Hüsken and P. Stagge, "Recurrent neural networks for time series classification," *Neurocomputing*, vol. 50, pp. 223–235, 2003.

[20] M. Van Heeswijk, Y. Miche, T. Lindh-Knuutila, P. A. Hilbers, T. Honkela, E. Oja, and A. Lendasse, "Adaptive ensemble models of extreme learning machines for time series prediction," in *Artificial Neural Networks–ICANN 2009*. Springer, 2009, pp. 305–314.

[21] B. Glocker, O. Pauly, E. Konukoglu, and A. Criminisi, "Joint classification-regression forests for spatially structured multi-object segmentation," in *Computer Vision–ECCV 2012*. Springer, 2012, pp. 870–881.

[22] D. Zhang, D. Shen, A. D. N. Initiative *et al.*, "Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in alzheimer's disease," *NeuroImage*, vol. 59, no. 2, pp. 895–907, 2012.

[23] O. Anava, E. Hazan, and A. Zeevi, "Online time series prediction with missing data," in *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*, 2015, pp. 2191–2199.

[24] J. Xu, M. van der Schaar, J. Liu, and H. Li, "Forecasting popularity of videos using social media," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 9, no. 2, pp. 330–343, 2015.

[25] A. Zubiaga, D. Spina, R. Martínez, and V. Fresno, "Real-time classification of twitter trends," *Journal of the Association for Information Science and Technology*, vol. 66, no. 3, pp. 462–473, 2015.

[26] K. Lee, D. Palsetia, R. Narayanan, M. M. A. Patwary, A. Agrawal, and A. Choudhary, "Twitter trending topic classification," in *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*. IEEE, 2011, pp. 251–258.

[27] R. A. Jacobs and J. K. Kruschke, "Bayesian learning theory applied to human cognition," *Wiley Interdisciplinary Reviews: Cognitive Science*, vol. 2, no. 1, pp. 8–21, 2011.

[28] J. B. Tenenbaum, T. L. Griffiths, and C. Kemp, "Theory-based bayesian models of inductive learning and reasoning," *Trends in cognitive sciences*, vol. 10, no. 7, pp. 309–318, 2006.

[29] G. Koop and D. Korobilis, *Bayesian multivariate time series methods for empirical macroeconomics*. Now Publishers Inc, 2010.

[30] D. Ray, B. King-Casas, P. R. Montague, and P. Dayan, "Bayesian model of behaviour in economic games," in *Advances in neural information processing systems*, 2009, pp. 1345–1352.

[31] S. Bikhchandani, D. Hirshleifer, and I. Welch, "A theory of fads, fashion, custom, and cultural change as informational cascades," *Journal of political Economy*, pp. 992–1026, 1992.

[32] A. Chakrabarty and X. Guo, "Optimal stopping times with different information levels and with time uncertainty," *Stochastic analysis and applications to finance*, vol. 13, pp. 19–38, 2007.

[33] D. Mokart, J. Lambert, D. Schnell, L. Fouché, A. Rabbat, A. Kouatchet, V. Lemiale, F. Vincent, E. Lengliné, F. Bruneel *et al.*, "Delayed intensive care unit admission is associated with increased mortality in patients with cancer with acute respiratory failure," *Leukemia & lymphoma*, vol. 54, no. 8, pp. 1724–1729, 2013.

[34] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," in *ICML*, vol. 3, 2003, pp. 856–863.

[35] L. Mardini, J. Lipes, and D. Jayaraman, "Adverse outcomes associated with delayed intensive care consultation in medical and surgical inpatients," *Journal of critical care*, vol. 27, no. 6, pp. 688–693, 2012.

[36] M. Cvach, "Monitor alarm fatigue: an integrative review," *Biomedical instrumentation and technology*, vol. 46, no. 4, pp. 268–277, 2012.

[37] J. P. Bliss and M. C. Dunn, "Behavioural implications of alarm mistrust as a function of task workload," *Ergonomics*, vol. 43, no. 9, pp. 1283–1300, 2010.

[38] J. P. Bliss, C. K. Fallon, and N. Nica, "The role of alarm signal duration as a cue for alarm validity," *Applied ergonomics*, vol. 38, no. 2, pp. 191–199, 2007.

[39] J. F. Dasta, T. P. McLaughlin, S. H. Mody, and C. T. Piech, "Daily cost of an intensive care unit day: the contribution of mechanical ventilation," *Clitical care medicine*, vol. 33, no. 6, pp. 1266–1271, 2005.

[40] A. Egol, R. Fromm, K. K. Guntupalli, M. Fitzpatrick, D. Kaufman, S. Nasraway, D. Ryon, and J. Zimmerman, "Guidelines for intensive care unit admission, discharge, and triage," *Critical Care Medicine*, vol. 27, no. 3, pp. 633–638, 1999.

[41] S. R. Bruijns, H. R. Guly, O. Bouamra, F. Lecky, and W. A. Lee, "The value of traditional vital signs, shock index, and age-based markers in predicting trauma mortality," *Journal of Trauma and Acute Care Surgery*, vol. 74, no. 6, pp. 1432–1437, 2013.

[42] A. Olaussen, T. Blackburn, B. Mitra, and M. Fitzgerald, "Review article: shock index for prediction of critical bleeding post-trauma: a systematic review," *Emergency medicine Australasia*, vol. 26, no. 3, pp. 223–228, 2014.

# VII. APPENDIX

## A. Proof of Theorem 1

In order to prove the Theorem, we first hold the following assumptions on the physiological stream stopping time and the patients' hospitalization time (time of admission to hospital), and hospitalization period (time between admission to hospital and transfer to ICU or discharge). We assume that the maximum hospitalization period for any patient is $\bar{T}_H$, the hospitalization time $t_H$ is random, and the stopping time $\tau_s$ is random where the distributions of hospitalization and stopping times are given by $f_{t_H}(t_H)$, $f_{\tau_s}(\tau_s \,|\mathcal{H}_0\,)$ and $f_{\tau_s}(\tau_s \,|\mathcal{H}_1\,)$, where $\text{supp}\left(f_{t_H}\left(t_H\right)\right) = [0, \bar{T}_H]$, $\text{supp}\left(f_{\tau_s}\left(\tau_s \,|t_H\,\right)\right) = [t_H, \bar{T}_H]$.

Let $B_t^*$ and $B_t$ be the belief processes of a truthful and a non-truthful belief systems respectively. A truthful belief system has access to the joint distributions of the physiological data stream $(\mathbb{P}_0, \mathbb{P}_1)$ and knows the stopping time $\tau_s$, whereas the non-truthful belief system maintains estimates of the joint distribution of the physiological data stream $(\mathbb{Q}_0, \mathbb{Q}_1)$, where $d\left(\mathbb{P}_m, \mathbb{Q}_m\right) > 0$ for a probability metric $d$. In the following, we show that both $B_t^*$ and $B_t$ are martingales with respect to the filtration $\mathcal{F}_t$. Note that

$$\begin{aligned} B_t^*\left(\mathcal{H}_1 \,|\mathcal{F}_t\,\right) &= \frac{\mathbb{P}\left(\{X_\tau\}_{\tau=t_H}^t \,|\mathcal{H}_1\right)\mathbb{P}\left(\mathcal{H}_1\right)}{\sum_{i\in\{0,1\}}\mathbb{P}\left(\{X_\tau\}_{\tau=t_H}^t \,|\mathcal{H}_i\right)\mathbb{P}\left(\mathcal{H}_i\right)} \\ &= \frac{B_{t-1}^*\left(\mathcal{H}_1 \,|\mathcal{F}_{t-1}\,\right)\mathbb{P}\left(X_t \,|\mathcal{H}_1\,\right)}{\sum_{i\in\{0,1\}} B_{t-1}^*\left(\mathcal{H}_i \,|\mathcal{F}_{t-1}\right)\mathbb{P}\left(X_t \,|\mathcal{H}_i\right)}. \end{aligned} \tag{10}$$

Thus, we have that

$$\begin{aligned} \mathbb{E}\left[B_{t+1}^* \,|\mathcal{F}_{t+1}\right] &= \mathbb{E}\left[\frac{B_t^*\left(\mathcal{H}_1 \,|\mathcal{F}_t\,\right)\mathbb{P}\left(X_t \,|\mathcal{H}_1\,\right)}{\sum_{i\in\{0,1\}} B_t^*\left(\mathcal{H}_i \,|\mathcal{F}_t\,\right)\mathbb{P}\left(X_t \,|\mathcal{H}_i\right)}\right] \\ &= \sum_{X_t\in\mathcal{X}_t}\frac{B_t^*\left(\mathcal{H}_1 \,|\mathcal{F}_t\,\right)\mathbb{P}\left(X_t \,|\mathcal{H}_1\,\right)\mathbb{P}(X_t)}{\sum_{i\in\{0,1\}} B_t^*\left(\mathcal{H}_i \,|\mathcal{F}_t\,\right)\mathbb{P}\left(X_t \,|\mathcal{H}_i\right)} \\ &= \sum_{X_t\in\mathcal{X}_t} B_t^*\left(\mathcal{H}_1 \,|\mathcal{F}_t\,\right)\mathbb{P}\left(X_t \,|\mathcal{H}_1\,\right) \\ &= B_t^*\left(\mathcal{H}_1 \,|\mathcal{F}_t\,\right)\sum_{X_t\in\mathcal{X}_t}\mathbb{P}\left(X_t \,|\mathcal{H}_1\,\right) \\ &= B_t^*\left(\mathcal{H}_1 \,|\mathcal{F}_t\,\right). \end{aligned} \tag{11}$$

Since $\mathbb{E}\left[B_{t+1}^* \,|\mathcal{F}_{t+1}\right] = B_t^*\left(\mathcal{H}_1 \,|\mathcal{F}_t\,\right)$, then the truthful belief process is martingale. Now we focus on the non-truthful belief process $B_t$, which we can write as

$$B_t\left(\mathcal{H}_1 \,|\mathcal{F}_t\,\right) = \frac{B_{t-1}\left(\mathcal{H}_1 \,|\mathcal{F}_{t-1}\,\right)\mathbb{Q}\left(X_t \,|\mathcal{H}_1\,\right)}{\sum_{i\in\{0,1\}} B_{t-1}\left(\mathcal{H}_i \,|\mathcal{F}_{t-1}\,\right)\mathbb{Q}\left(X_t \,|\mathcal{H}_i\right)}. \tag{12}$$

Thus, we have that

$$\begin{aligned} \mathbb{E}\left[B_{t+1} \,|\mathcal{F}_{t+1}\right] &= \mathbb{E}\left[\frac{B_t\left(\mathcal{H}_1 \,|\mathcal{F}_t\,\right)\mathbb{Q}\left(X_t \,|\mathcal{H}_1\,\right)}{\sum_{i\in\{0,1\}} B_t\left(\mathcal{H}_i \,|\mathcal{F}_t\,\right)\mathbb{Q}\left(X_t \,|\mathcal{H}_i\right)}\right] \\ &= \sum_{X_t\in\mathcal{X}_t}\frac{B_t\left(\mathcal{H}_1 \,|\mathcal{F}_t\,\right)\mathbb{Q}\left(X_t \,|\mathcal{H}_1\,\right)\mathbb{P}(X_t)}{\sum_{i\in\{0,1\}} B_t\left(\mathcal{H}_i \,|\mathcal{F}_t\,\right)\mathbb{Q}\left(X_t \,|\mathcal{H}_i\right)} \\ &= B_t\left(\mathcal{H}_1 \,|\mathcal{F}_t\,\right). \end{aligned} \tag{13}$$

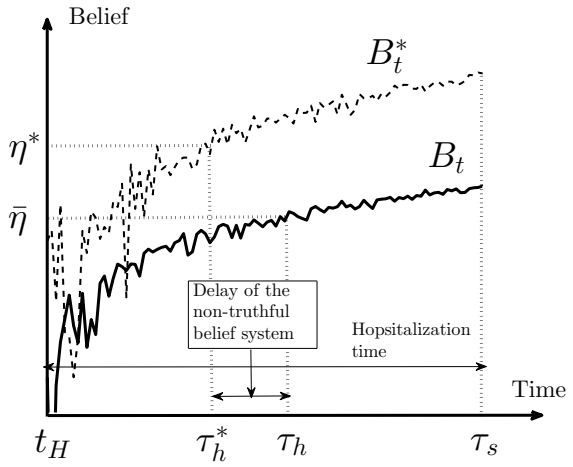


Fig. 4. Depiction for the belief process as computed by a truthful and a non-truthful belief systems.

Now define the threshold type strategies $\eta^*$ (a threshold on $B_t^*$) and $\bar{\eta}$ (a threshold on $B_t$) as follows:

$$\eta^* = \arg \sup_{\eta \in [0,1]} \mathbb{E}_{\mathbb{P}} \left[ g\left( \{X_\tau\}_{\tau=0}^{\tau_h(\eta)} \right) \mathbf{1}_{\{\tau_h(\eta) < \tau_s\}} \right], \tag{14}$$

and

$$\bar{\eta} = \arg \sup_{\eta \in [0,1]} \mathbb{E}_{\mathbb{Q}} \left[ g\left( \{X_\tau\}_{\tau=0}^{\tau_h(\eta)} \right) \mathbf{1}_{\{\tau_h(\eta) < \tau_s\}} \right]. \tag{15}$$

As shown in Fig. 4, the non-truthfulness of the belief system may lead for instance to a delay in the ICU alarm. Now we focus on a certain realization of the stopping time $\tau_s$. Since $\sup_t B_t = \sup_t B_t^* = 1$, and $\inf_t B_t = \inf_t B_t^* = 0$, then $\mathbb{E}[B_t] < \infty$ and $\mathbb{E}[B_t^*] < \infty$, i.e. $B_t$ and $B_t^*$ are bounded martingales. Thus, by Doob's martingale convergence theorem, we know that $B_t \to B_\infty$ and $B_t^* \to B_\infty^*$ almost surely, where $\mathbb{E}[B_\infty^*] < \infty$, and $\mathbb{E}[B_\infty] < \infty$. It is easy to show that the sequence $B_t^* - B_t$ is also a martingale with respect to the filtration $\mathcal{F}_t$, i.e. $\mathbb{E}[B_t^* + 1 - B_t + 1 \,|\mathcal{F}_t] = B_t^* - B_t$. Now recall that we want to show that $\mathbb{P}(|V^* - V(\bar{\eta})| < \epsilon) > 1 - \delta$. To prove this, it suffices to show that there exists $\epsilon^{'} \in [0,1]$, such that $\mathbb{P}\left(|\eta^* - \bar{\eta}| < \epsilon^{'}\right) > 1 - \delta$. This is equivalent to show that the martingale sequence $B_t^* - B_t$ converges to a value less than $\epsilon^{''} \in [0,1]$ with a probability $1 - \delta$. This is satisfied if for $N^*(\epsilon, \delta)$, there exists an algorithm $\mathcal{A}^D$ that if used to estimate $\mathbb{Q}$, it will prompt a distribution that is within a Kolmogorov-Smirnov distance of $\Delta(\epsilon)$ from the true distribution $\mathbb{P}$. By Dvoretzky-Kiefer-Wolfowitz inequality, we know that if the algorithm $\mathcal{A}^D$ just computes $\mathbb{Q}$ as the empirical distribution, then we have that

$$\Pr\left( \sup_{t \in [t_H, t_H + \tau_s]} \left| \mathbb{Q}_m^t - \mathbb{P}_m^t \right| > \Delta(\epsilon) \right) \leq 2\exp\left(-2N\Delta^2(\epsilon)\right).$$

Thus, we can find $N^*(\epsilon, \delta)$ by equating $1 - \delta$ with the RHS in the equation above, and for any $N > N^*(\epsilon, \delta)$, we have that $\mathbb{P}(|V^* - V(\bar{\eta})| < \epsilon) > 1 - \delta$.

*B. Pseudo-code of ForecastICU*

**Offline Stage**:

*Input*: $\mathbf{X}_0^{ref}, \mathbf{X}_1^{ref}, T_s$

**1) Data Reconstruction**

**for** i = 1 **to** N **do**

$\tilde{\mathbf{X}}_{(i)}^{ref} = h_{spline}(\{\mathbf{X}_{(i)}^{ref}(m,n)\}_{n=0}^{K-1}, T_s)$

**end for**

**2) Relevant Feature Selection**

$\tilde{\mathbf{Y}}^{ref} = CFS(\tilde{\mathbf{X}}^{ref})$,

**3) Parametric density estimation**

$[\hat{\mu}_m^t(j)]_{j=1}^R = \frac{1}{N_m}\sum_{i=1}^{N_m} \tilde{\mathbf{Y}}_{(i),m}^{ref}(j,t)$

$[\hat{\mathbf{\Sigma}}_m^t]_{k,l} = \frac{1}{N_m-1}\sum_{i=1}^{N_m} \bar{\mathbf{Y}}_{(i),m}^{ref}(k,t)\bar{\mathbf{Y}}_{(i),m}^{ref}(l,t)$

**Real-time Stage**:

*Input*: $\{\mathbf{X}_\tau\}_{\tau=0}^{T^H}, \gamma, \eta, W$

**for** t = 1 **to** $T^H$ **do**

**1) Current State Estimation**

**for** m = 0 **to** 1 **do**

$T_m^*(t) = \arg\max_\tau \mathbb{Q}(\{X_k\}_{k=\tau-t+1}^{\tau}|\mathcal{H}_m)$

**end for**

**2) Belief Update Algorithm**

$B_t(\mathcal{H}_1|\mathcal{F}_t) = \mathbb{Q}(\mathcal{H}_1|\{X_\tau\}_{\tau=t_o}^t)$

$= \frac{N_1\mathbb{Q}(\{X_\tau\}_{\tau=t_o^1}^t|\mathcal{H}_1, T_1^*(t))}{N_0\mathbb{Q}(\{X_\tau\}_{\tau=t_o^0}^t|\mathcal{H}_0, T_0^*(t)) + N_1\mathbb{Q}(\{X_\tau\}_{\tau=t_o^1}^t|\mathcal{H}_1, T_1^*(t))}$

$\tilde{B}_t(\mathcal{H}_1|\mathbb{F}_t) = \frac{1}{W}\sum_{\tau=t-W}^t B_\tau(\mathcal{H}_1|\mathcal{F}_\tau)$

**3) Sequential Decision Making**

$$\text{Decision}(t) = \begin{cases} \mathcal{H}_1 & \text{if } \hat{B}_t(\mathcal{H}_1|\mathcal{F}_t) \geq \eta \\ \mathcal{H}_0 & \text{otherwise} \end{cases}$$

**end for**

Fig. 5. Pseudo-code of ForecastICU

### C. Features of the Dataset

TABLE III
ENTIRE FEATURE INFORMATION

| NO | FEATURE NAME |
|---|---|
| | **Time Dependent Continuous Features** |
| 1 | SYSTOLIC BLOOD PRESSURE |
| 2 | DIASTOLIC BLOOD PRESSURE |
| 3 | HEART RATE |
| 4 | RESPIRATORY RATE |
| 5 | TEMPERATURE |
| 6 | O2 SATURATION |
| 7 | WHITE BLOOD CELL |
| 8 | HEMOGLOBIN |
| 9 | PLATELET COUNT |
| 10 | SODIUM |
| 11 | POTASSIUM |
| 12 | CHLORINE |
| 13 | CO2 SATURATION |
| 14 | BLOOD UREA NITROGEN |
| 15 | CREATINE |
| 16 | GLUCOSE |
| | **Time Dependent Discrete Features** |
| 17 | O2 DEVICE (BINARY) |
| 18 | BREATH ASSIST DEVICE (49 CATEGORIES) |
| | **Time Independent Features** |
| 19 | AGE |
| 20 | ETHNICITY |
| 21 | RACE |
| 22 | GENDER |
| 23 | TRANSFER (BINARY) |
| 24 | ADMITTED SOURCE |

*1) Entire feature information:*

TABLE IV
RELEVANT FEATURES FOR ICU ADMISSION PREDICTION

| Rank | Acronym | Relevant Features |
|---|---|---|
| 1 | RR | Respiratory Rate |
| 2 | HR | Heart Rate |
| 3 | BUN | Blood Urea Nitrogen |
| 4 | GLU | Glucose |
| 5 | Breath | Oxygen Supply Device (Binary) |
| 6 | DBP | Diastolic Blood Pressure |
| 7 | SPO2 | O2 Saturation |

*2) Relevant Features for ICU Admission Prediction:* Based on the correlation feature selection (CFS) algorithm with minimum redundancy and maximum relevance (mRMR) criterion, we discover 7 relevant features among the entire 24 features which are highly correlated with ICU admission but poorly correlated with each other. Table IV explicitly lists 7 relevant features and these can be justified by the medical references [40] [41] [42]. Note that all of the relevant features are time dependent features.

### D. Model Justifications

*1) Martingale Properties:* In the paper, we assume that physiological data streams of ICU and DIS patients can be modeled as stochastic processes with sub/supermartingales and martingales properties, respectively. Fig. 6 illustrates four representative physiological data streams of ICU patient that tend to increases/decreases when the actual ICU event time approaches (sub/supermartingales). On the other hand, the representative physiological data streams of DIS patient are consistent within the entire hospitalization periods (martingales).

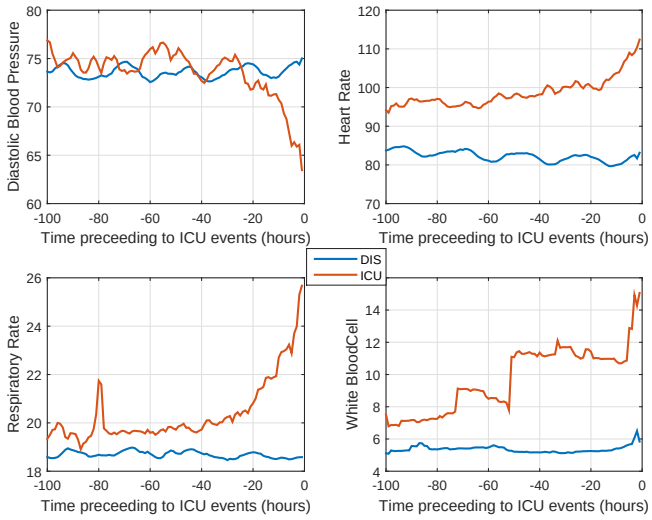


Fig. 6. The average temporal physiological data streams

*2) Multivariate Gaussian Distribution Approximation:* In this paper, we also assume that the joint distribution of the physiological data streams can be modeled as a Multivariate Gaussian process. Fig. 7 illustrates the histogram of the systolic blood pressure and heart rate extracted by the reconstructed dataset of ICU and DIS patients, respectively. As it can be seen, these can be indeed modeled as Gaussian distributions - the fitting error is less than 10%. Fig. 8 shows that the joint distributions between the physiological features can indeed be modeled using a Multivariate Gaussian distribution..

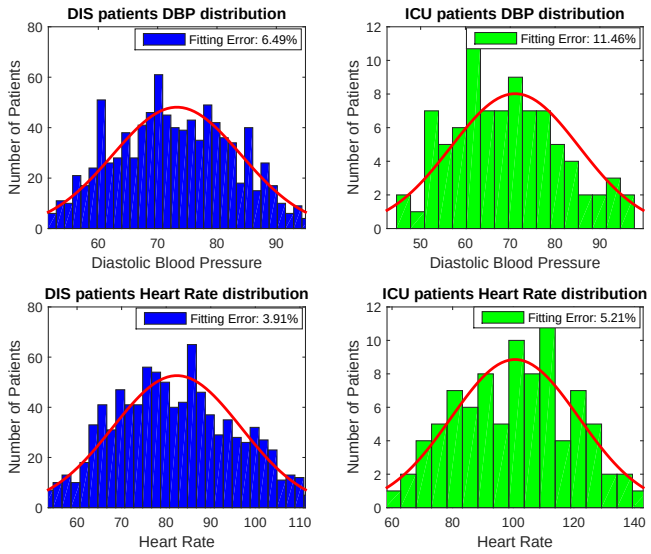


Fig. 7. Histograms of diastolic blood pressures and heart rates at 10 hours before ICU/DIS events.

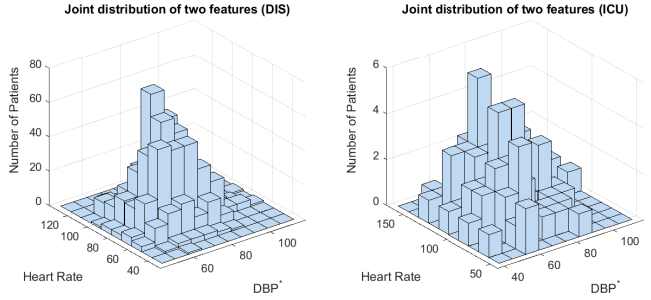


Fig. 8. Joint distribution of diastolic blood pressure and heart rates

### *E. Extension of ForecastICU: Patient Risks Tracking Systems (PRTS)*

ForecastICU can be extended to patients risks tracking systems (PRTS) which keeps tracking the ICU belief (risks of ICU admission) until the actual ICU admission or discharge event. This system is useful in real clinical setting because PRTS helps doctors to focus on the real-time high risk patients based on the ICU belief provided by the algorithm. In this subsection, we illustrate the performance of ForecastICU in PRTS setting.

ForcastICU has a consistently higher PPV in comparison to other benchmarks which is represented in Table V and Fig. 9. For instance, given 70% TPR, ForecastICU achieves 80.1% PPV which is 5.2% better than the second best algorithm (Lasso Regularization). Moreover, with 70% PPV, Forecast ICU achieves 78.0% TPR which is 4.7% better than the second best algorithm. AUC values are also 1.5% higher than the second best algorithm and the p-value of the hypothesis test comparing ForecastICU and the second best algorithm is $\leq 0.01$.

TABLE V

PERFORMANCE COMPARISON OF ICU PREDICTION IN PRTS SETTING

| **Algorithms** | **TPR(%)** | **PPV(%)** |
|---|---|---|
| ForecastICU | 70.3± 1.75% | 80.1± 1.23% |
| Logistic Regression | 70.5± 1.13% | 73.5± 2.09% |
| Lasso Regularization | 70.1± 1.49% | 74.9± 1.98% |
| Random Forest | 70.7± 1.34% | 56.1± 1.24% |
| SVMs | 70.0± 1.28% | 44.9± 1.74% |

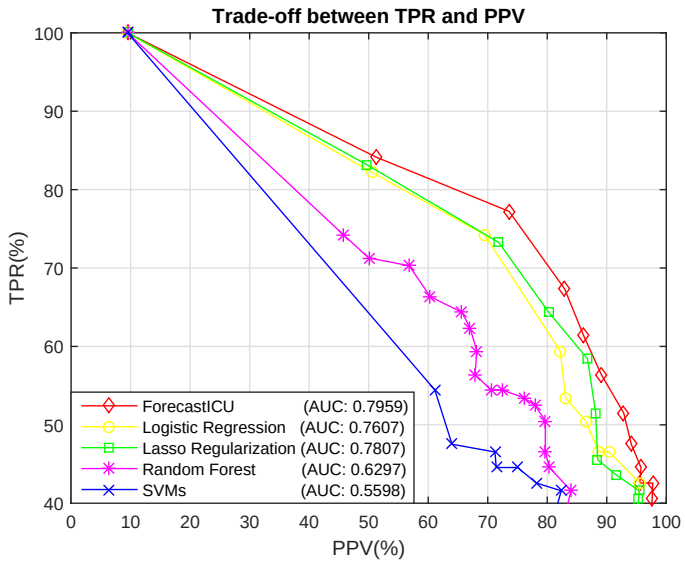


Fig. 9. Trade-off between TPR and PPV in PRTS setting

### F. Additional Experiment Results

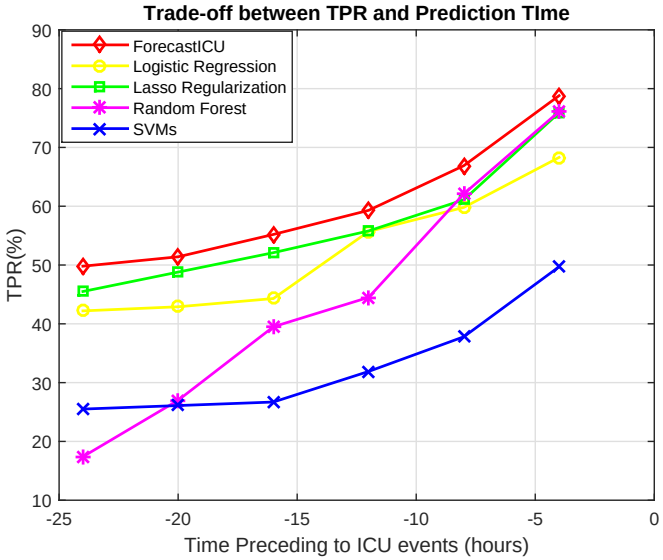


Fig. 10. Trade-off between TPR and the prediction time (fix PPV 30%)