Check for updates

Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
0(0) 1–15
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0962280220929045
journals.sagepub.com/home/smm
SAGE

Utilizing time series data embedded in electronic health records to develop continuous mortality risk prediction models using hidden Markov models: A sepsis case study

Akash Gupta[1], Tieming Liu[2] and Christopher Crick[2]

Abstract

Continuous mortality risk monitoring is instrumental to manage a patient's care and to efficiently utilize the limited hospital resources. Due to incompleteness and irregularities of electronic health records (EHR), developing continuous mortality risk prediction using EHR data is a challenge. In this study, we propose a framework to continuously monitor mortality risk, and apply it to the real-world EHR data. The proposed method employs hidden Markov models (temporal technique) that take account of both the previous state of patient's health and the current value of clinical signs. Following the Sepsis-3 definition, we selected 3898 encounters of patients with suspected infection to compare the performance of temporal and non-temporal methods (Decision Tree (DT), Logistic Regression (LR), Naive Bayes (NB), Random Forest (RF), and Support Vector Machine (SVM)). The area under receiver operating characteristics (AUROC) curve, sensitivity, specificity and G-mean were used as performance measures. On the selected data, the AUROC of the proposed temporal framework (0.87) is 9–12% greater than the non-temporal methods (DT: 0.78, NB: 0.79, SVM: 0.79, LR: 0.80 and RF: 0.80). The results also show that our model (G-mean$^1/_4$0.78) provides a better balance between sensitivity and specificity compared to clinically acceptable bed-side criteria (G-mean$^1/_4$0.71). The proposed framework leverages the longitudinal data available in EHR and performs better than the non-temporal methods. The proposed method facilitates information related to the time of change of the patient's health that may help practitioners to plan early and develop effective treatment strategies.

Keywords

Statistical models, time series model, decision support system, hidden Markov model, infection, sepsis

1 Introduction

The continuous mortality risk prediction models are key to early diagnosis and to composing efficient treatment strategies. The continuous mortality risk prediction models assess the risk of mortality throughout a patient's hospital stay using longitudinal observed clinical signs such as heart rate, systolic blood pressure, etc. The risk of mortality varies over time; therefore, continuous mortality risk prediction models provide relevant temporal

[1]California State University, Northridge, Northridge, CA, USA
[2]Oklahoma State University, Stillwater, Stillwater, OK, USA

Corresponding author:
Akash Gupta, California State University, Northridge, BB 4124, Northridge 91330-0001, CA, USA.
Email: akash.gupta@csun.edu

insights such as time of change in patients' health and time duration for a disease being in a specific state. In this paper, we use term *mortality progression* to present *continuous mortality risk prediction*.

The current state of the art mortality prediction models use aggregated (mean, median or peak) clinical measurements.[1,2] Goldstein et al. showed in their comprehensive review paper on a risk prediction model using electronic health record (EHR) that 93% of studies do not leverage longitudinal information present in EHR data.[3] The longitudinal data capture important variations in clinical signs and can be used to develop continuous mortality progression models. From 2009 to 2014, the number of hospitals utilizing a basic EHR system increased from 12.2% to 75.5%,[4] and EHR data present its own challenges: sparsity of lab measurements, missing clinical measurements and unrecorded intermittent severity of disease during the hospital stay. Therefore, designing a framework to develop continuous mortality prediction models using EHR is meaningful.

In this study, we utilize the longitudinal data and propose a hidden Markov model (HMM) framework to model mortality progression. HMM is a probabilistic graphical technique to model time series problems. The proposed framework computes mortality risk in combination with present observations and past trends of clinical signs. This modeling approach is closely aligned with the real-world phenomenon where physicians use historical data in addition to current vital signs to assess the health of patients.

The patients with suspected infection can lead to sepsis if not treated timely. Sepsis is a *life-threatening organ dysfunction caused by a dysregulated host response to infection*.[5] Sepsis has a high in-hospital mortality rate[6] and an extremely high increase in incidence rate.[7] At about 5.2% of national hospital costs (or $20 billion), sepsis is considered the most expensive condition treated in U.S. hospitals.[8] Due to increasing concerns related to sepsis diagnosis and treatment, beginning from July 25, 2018, the Center for Medicare & Medicaid Services (CMS) has started releasing data on hospital performance on sepsis care. This publicly available data will push hospitals to take steps towards improving sepsis care. One of the possible ways of improving sepsis care is to integrate a mortality monitoring system in existing hospital data architecture to understand the progression of infection among suspected infection patients. The progression of the infection and mortality is positively related.[9] In this study, severe infection also referred to as sepsis. We selected a population of patients with suspected infection to show the meaningful application of the proposed framework. To be noted that in our context, monitoring signifies the understanding of the severity of the disease, not the quality of the care.

We propose a hidden Markov model framework that combines multiple time series data of clinical variables and generates inference on mortality progression. The performance of the proposed model is compared to the performance of popularly employed traditional predictive modeling techniques (decision tree, logistic regression, naive Bayes, random forest and support vector machine). We call such traditional methods *non-temporal* as they do not require time series data. In contrast, our proposed framework leverages available time series data in EHR; therefore, we classify it as *temporal*.

This study:

1. Proposes a HMM framework to develop mortality progression using EHR data.
2. Underlines the advantages of using temporal methods over non-temporal techniques.
3. Extracts relevant time information that can help physicians to compose better treatment strategies and efficiently utilize hospital resources.
4. Develops a mortality progression model to understand the progression of infection using only non-invasive clinical signs.

2 Related work

2.1 Non-temporal methods for mortality prediction

The use of non-temporal machine learning and statistical methods in combination with aggregated value (mean, median, or peak) of clinical signs is growing. Matheny et al. used peak or mean value (depending on the type of lab) to architect a risk stratification model for hospital-acquired acute kidney injury using logistic regression.[1] Saltzman et al. developed an in-hospital mortality prediction model using decision trees.[10] The authors used clinical measurements recorded at the time of admission into an emergency department. Ramchandran et al. proposed a mortality prediction model for cancer patients using logistic regression.[11] To implement the model, the authors used measurements recorded within the first 24 hours of the admission. Tabak et al., instead of developing a specific disease mortality model, developed a generic mortality prediction model using 23 lab results and 2 demographic measurements.[2] The authors developed logistic model using the first measurement of hospital visits.

All of these proposed models use either aggregated values or a specific measurement (first or last observation during hospital visit) for model implementation. Due to the use of a single value, these models do not capture the dynamic behavior of disease.

In addition to aforementioned studies, a few diagnostic criteria are established to capture the progression of infection: Systemic Inflammatory Response Syndrome (SIRS),[12] quick Sepsis-related Organ Failure Assessment (qSOFA),[5] Modified Early Warning System (MEWS),[13] and Sepsis-related Organ Failure Assessment (SOFA).[14] Theses diagnostic criteria calibrate the risk of mortality. SIRS and qSOFA are designed for early bed-side diagnosis,[15] while SOFA and MEWS are designed for understanding the criticality of patients in intensive care.[16] All of these criteria lack in taking account the variability of clinical variables to assess the risk of mortality.

2.2 Temporal methods for disease risk progression

Some studies are performed to understand the dynamic behavior of diseases. Henry et al. used Cox proportional hazard modeling techniques to develop the septic shock progression model. This study developed a targeted real-time warning score (TREW) to predict septic shock in the early stages.[17] Although authors utilized longitudinal data using proportional hazards model, this modeling technique does not incorporate past information while estimating the mortality risk. The principal of a proportional hazards model lies in the number of subjects that survive over the study time window.

Van Wyk et al. proposed a model to detect the onset of sepsis by using random forest with the continuous stream of the clinical variables.[18] Although the proposed model used clinical measurements over different time intervals to generate the prediction, however, the model ignores the past state of the patient's criticality to establish the prediction. Peelen et al. investigated the sequence of organ failure for patients in intensive care using dynamic Bayesian networks.[19] Cai et al. developed the non-disease-specific progression model using Bayesian networks.[20]

Some studies developed clinical decision support systems using neural networks and EHR data.[21,22] Raghu et al. developed a treatment policy recommendation system for sepsis patients in intensive care.[23] The proposed method employed reinforcement learning on the time-series clinical data available in EHR. Kam and Kim developed an early sepsis detection model using deep learning methods.[24] The model predicts mortality based on the temporal clinical measurements. The most significant limitation of the applicability of deep learning methods is the lack of interpretation. The understanding of the outcome is the critical factor for successful implementation of clinical models. Hence, we avoided the use of neural networks-based methods.

HMMs enable modeling of the longitudinal behavior of a disease with interpretation. A few researchers have applied HMMs to model the criticality of diseases. Liu et al. proposed a 2 D-continuous time HMM to model glaucoma progression using longitudinal data.[25] Vairavan et al. proposed a mortality prediction model for intensive care using the combination of HMM and logistic regression.[26] In the proposed method, the author used the output of an HMM model as input to the logistic regression. This study does not capture the trajectory of risk assessment. The online risk assessment enables time-relevant information such as the time at which a patient transit to high risk to low risk or vice-versa, and duration of how long a patient is in a specific state of disease.

Sukkar et al. developed a slow progressing disease risk model using HMMs.[27] This work set the foundation of our work where we propose a mortality progression model rather than disease progression using EHR data. The primary difference is that Sukkar et al. did not test the proposed framework on EHR data. The study collected data from brain images. With the rapid growth of EHR data, developing the framework to establish a mortality progression model using EHR is instrumental. The secondary difference is that the author did not comment on the significance of using an HMM model over non-temporal techniques. In this study, we bridge the gap by proposing a framework that can model mortality progression and can be easily replicated for any disease.

3 Method

3.1 Proposed framework

The proposed framework to develop a mortality monitoring system using an HMM is shown in Figure 1. The framework includes three steps: data preprocessing, training HMM parameters and inference. EHR data are stored in long form where each clinical measurement of a patient's hospital stay is stored in separate rows. The proposed framework requires data to be in wide form where longitudinal measurements are stored in one row. The data preprocessing steps are explained in detail in the next section. The preprocessed data is used to learn the

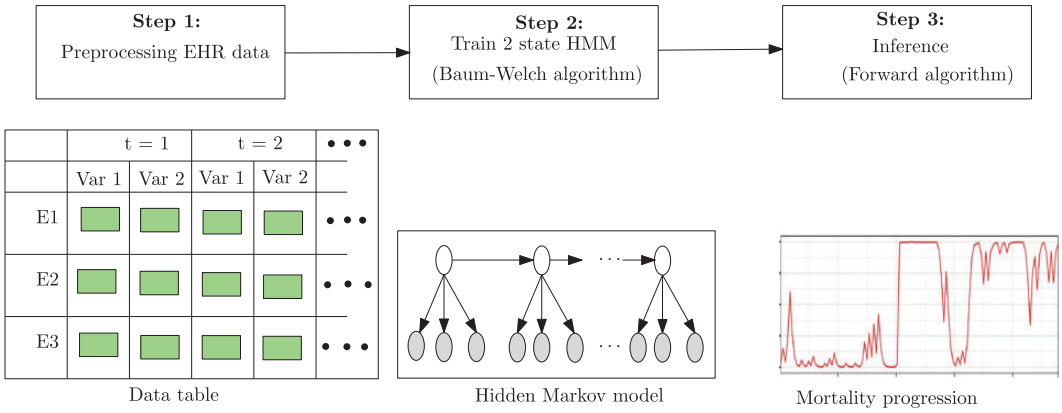

Figure 1. Proposed framework. E1 stands for Encounter 1. Similarly, E2, E3, etc.

two-state HMM parameters using the *Baum-Welch algorithm*. The two states represent stable and critical patient status. A brief explanation of HMMs and their training is provided in Sections 3.3 and 3.4, respectively. The learned model combined with longitudinal observed clinical measurements is used for generating the inference for mortality progression using the *forward algorithm* (explained in Section 3.5).

3.2 EHR data extraction and preprocessing

Although this section explains the data extraction and preprocessing steps for EHR data used in this study to perform the numerical experiments, the procedure is generic and could be applied to any EHR dataset. The study data was collected from Cerner Corporation's Health Facts warehouse, one of the largest de-identified and HIPAA compliant health care data repositories in the United States. The data included longitudinal date- and time-stamped information on admission and discharge, laboratory results, diagnosis code, patient demographics, and additional clinical and billing information.

The selection of population was derived following the Sepsis-3 definition.[9] The gold standard constitutes a combination of culture drawn (Event 1) and antibiotics administration (Event 2) within a specific time interval. There are two scenarios: (1) Event 1 occurs first, then Event 2 occurs within 72 hours (N = 52246), (2) Event 2 occurs first, then Event 1 occurs within 24 hours (N = 196543). The extracted data includes patient visits that satisfy either of two aforementioned scenarios. The time of infection is the time of occurrence of the first event. The selection of antibiotic and culture was based on the consultation with knowledge experts.

The data extraction steps are illustrated in Figure 2(a). The raw data include over 60 million patient visits from about 600 hospitals. From the raw dataset, we extracted information about the population of interest: patients with suspicion of infection. The total patient visits with suspicion of infection were 248,789. From this population, we selected patient visits that could be used for the modeling. The final dataset used for modeling included 3898 patient visits. The significant drop in data size is due to the poor quality of the medical data. In addition, we considered patient visits that resulted in the same length of stay to explain the importance of utilizing the longitudinal data available in EHR at the same time keeping the computational complexity at the minimum. In this study, we considered encounter visits that resulted in discharge on the eleventh day. For such encounters we only used the data till tenth day for training the model. The parameters learned using such data perform equally well for the data of any length because the estimated probability of a hidden state in HMM only depends on the immediate past not the length of the observed sequence. For the interested reader, we included the proportion of discharge types for each day in Appendix A Supplementary Material.

Figure 2(b) shows the preprocessing steps of data preparation. The emphasis of the study is not to develop a new predictive model but to underline the benefits of applying the proposed temporal framework over traditional non-temporal methods. Therefore, we borrowed the set of three non-invasive clinically useful variables (systolic blood pressure, Glasgow Coma Scale score and respiratory rate) identified by the Sepsis-3 study[9] to capture the progression of infection. The Sepsis-3 study had shown, with comprehensive experiments, that systolic blood pressure, Glasgow Coma Scale score and respiratory rate are the key identifiers among suspected infection patients to understand the progression of the infection that potentially lead to sepsis.[9] Therefore, we limited

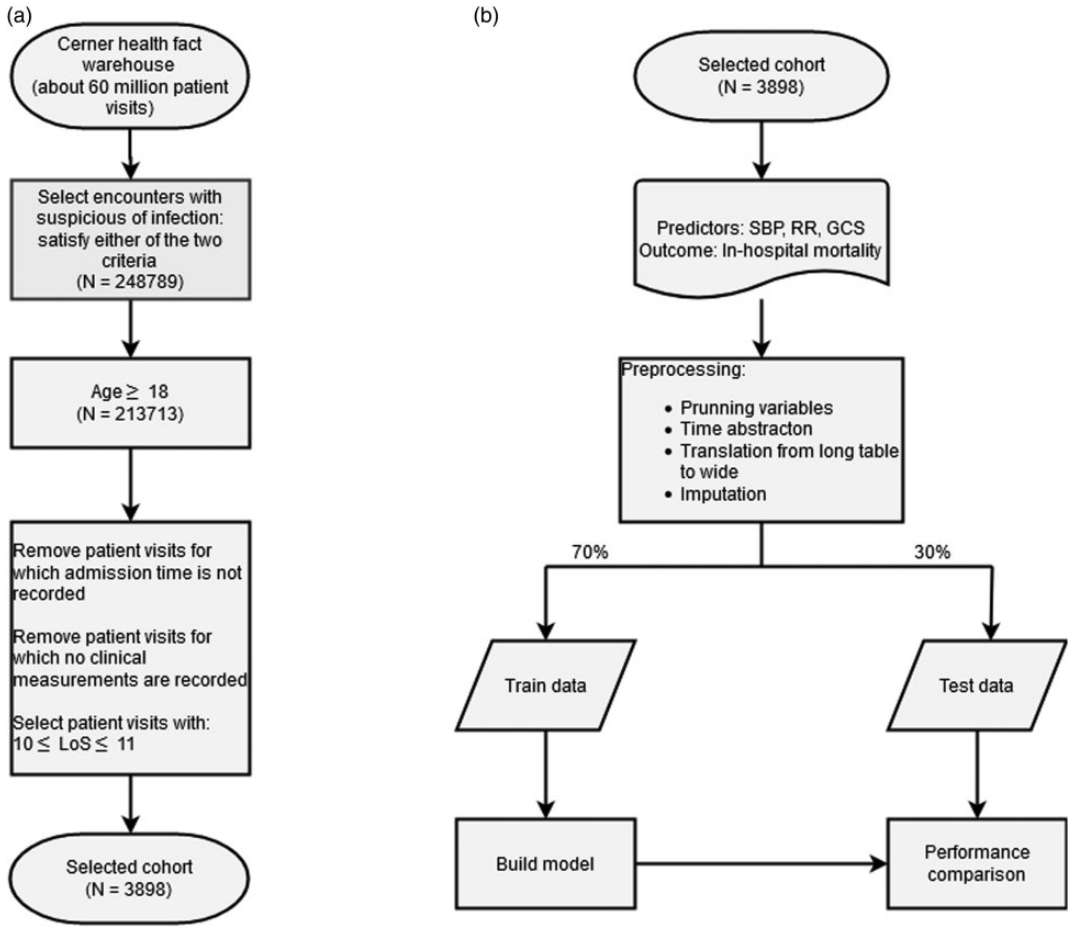

Figure 2. Data extraction (a) and data preprocessing (b). SBP: Systolic blood pressure, RR: respiratory rate, GCS: Glasgow Coma Scale score.

our study to only these routinely available vital signs because such clinical variables are more frequently recorded in EHR than laboratory tests and present comparatively fewer missing longitudinal data.

The EHR data may suffer many challenges. For example, clinical variables are stored with multiple names and varying units, and some measurements of clinical signs are meaningless. To ameliorate the inconsistencies, we preprocessed the data. The pruning of a variable includes removing infeasible clinical observations that may be present due to errors. For example, a respiratory rate of -36 is meaningless, and therefore, we removed such measurements. The extreme value proportion for each variable is listed in Appendix B, Supplementary Material.

Time abstraction is the process of aggregating the clinical measurements over time into time windows. Figure 3 graphically represents the time abstraction. In EHR data, a few clinical signs were recorded in intervals of 15 minutes, and some were a couple of hours apart. We aggregated the data into 2-h blocks to maintain the original variations in clinical signs and at the same time to reduce the volume of missing records. The selection of 2 h time window is also motivated by the recommended reassessment frequency of vital signs.[28] As shown in Figure 3, if more than one result is present in a 2-h interval, the mean of all results was used for analysis. Imputation is a process of finding the best fit measurement for missing values. The *left-center-right* technique was employed that is discussed in Gupta et al..[29] The preprocessed data was split into training (70%) and testing (30%) for model building and testing the performance, respectively.

3.3 Hidden Markov model

HMMs are popularly known for successfully modeling time series problems such as speech recognition. HMM consists of observed states and hidden (unknown) states. Unlike Markov models, in HMM, the state of interest is unobservable. Therefore, we call such modeling methods HMM. The aim of HMMs is to infer about the hidden

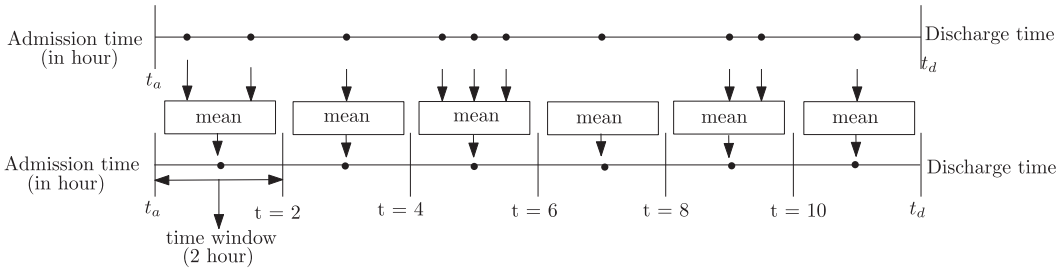

Figure 3. Time abstraction (t_a: admission time, t_d: discharge time).

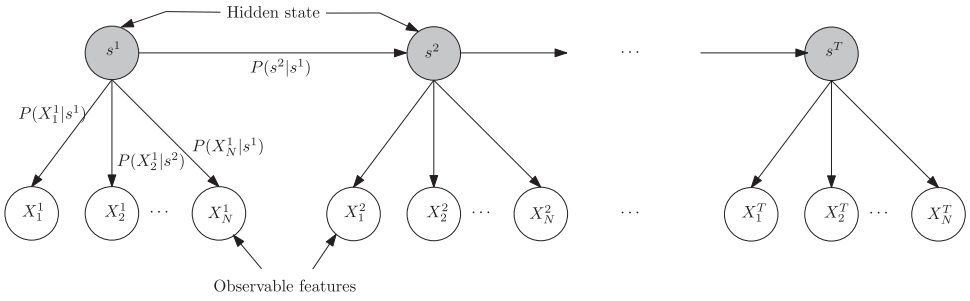

Figure 4. Hidden Markov model.

state of the system using the observable time series data. The basic framework of an HMM is shown in Figure 4. The shaded circles represent the hidden variables and unshaded circles represent observed variables. The arcs represent conditional probabilities. The conditional probabilities are divided into two classes:

1. *Transition probabilities*
   The transition probabilities reflect the probability of transition from one state at time $t-1$ to another state at time t

$$P(s^t|s^{t-1})$$

2. *Emission probabilities*
   The emission probabilities represent the probability of emitting an observation given the state.

$$P(X_i^t|s^t)$$

We assume that the parameters are time invariate. Therefore, $P(X_i^1|s^1) = P(X_i^2|s^2) = P(X_i^t|s^t) = P(X_i|s)$ and $P(s^2|s^1) = P(s^3|s^2) = P(s^T|s^{T-1})$

Here $X_i^1, X_i^2, \ldots, X_i^T$ represent the sequence of observed measurements for X_i ($i \in \{1, 2, \ldots, \mathrm{N}\}$), where T is the number of time steps and N is the number of observed variables. We assume that given the state at time t, the observed variables $(X_1^t, X_2^t, \ldots, X_N^t)$ are independent from each other. This assumption is in line with the basic principle of non-temporal methods such as logistic regression and naive Bayes.

$$P(X_1^t, X_2^t, \ldots, X_N^t|s^t) = \prod_{i=1}^{N} P(X_i^t|s^t) \tag{1}$$

As mentioned earlier, each sequence could be of a different length but to maintain the simplicity, we considered sequences of the same length. s^t denotes the state of the disease (1 = non-critical or 2 = critical) at time t ($t \in \{1, 2, \ldots, \mathrm{T}\}$).

3.4 Learning HMM parameters

The HMM parameters (*emission* and *transition* probabilities) are estimated using the maximum likelihood method. The objective is to maximize the likelihood of sequences of observations given the parameter as shown in equation (2)

$$\log L = \sum_{i=1}^{M} \log P(\mathbf{Z}_i|\lambda) \tag{2}$$

where Z_i is ith sequence of observations, $\mathbf{Z_i} = [(X_1^1, X_2^1, \ldots, X_N^1), (X_1^2, X_2^2, \ldots, X_N^2), \ldots, (X_1^t, X_2^t, \ldots, X_N^t), \ldots, (X_1^T, X_2^T, \ldots, X_N^T)]_i$. λ is the collection of HMM parameters of both transition and emission probabilities. Each sequence of observation is obtained from an independent hospital stay. M is the number of independent observed sequences corresponding to the number of patient's hospital stays (or widely known as *encounters*) in the dataset. For simplicity of notation, let $(X_1^t, X_2^t, \ldots, X_N^t)$ be $\mathbf{X^t}$, therefore, $\mathbf{Z} = [\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^t, \ldots, \mathbf{X}^T]$. We used *Baum-Welch algorithm* to estimate the HMM parameters such that the likelihood of observed data could be maximized.[30]

3.5 HMM inference

HMM inference includes computation of the most probable sequence of states. The most probable sequence of state is derived by estimating the probability of being in state s at time t, given the observed data until t and HMM parameters ($P^t(s|\mathbf{X^1}, \mathbf{X^2}, \ldots, \mathbf{X^t}, \lambda)$). The inference is computed using a recursive procedure partially adapted from Rabiner.[30] Following naive Bayes principle, the recursive procedure is modified to accommodate multiple features instead of considering only one feature at a time. We used multiplication of conditional probabilities $\left(\prod_{i=1}^{N} P(X_i^t|s^t)\right)$ to account for the effect of N variables in state probabilities.

Let $\alpha^t(s)$ be the forward variable and is defined as joint probability of partial observed sequence until time t $(\mathbf{X^1}, \mathbf{X^2}, \ldots, \mathbf{X^t})$ and health being in state s, given HMM parameters.

$$\alpha^t(s) = P^t(\mathbf{X^1}, \mathbf{X^2}, \ldots, \mathbf{X^t}, s^t = s|\lambda)$$

The recursive procedure to compute $\alpha^t(s)$ is explained as follows:

1. *Initialization:* Initialize the joint probability of initial observation $\mathbf{X^1}$ and state s

$$\alpha^1(s) = \alpha^0(s)\prod_{i=1}^{N} P(X_i^1|s^1 = s, \lambda) \quad \text{for} \quad s \in \{1, 2\}$$

where $\alpha^0(s)$ is the initial probability of health being in state s when no observation was observed.
2. *Induction:* for t = 2 to T−1

$$\alpha^t(s) = \left[\sum_{i\in\{1,2\}} \alpha^{t-1}(i)P(s^t = s|s^{t-1} = i)\right] \times \prod_{i=1}^{N} P(X_i^t|s^t = s) \quad \text{for} \quad s \in \{1, 2\}$$

This step shows how a patient's health transitions to one of two states at time t from previous state at time $t-1$. $\alpha^{t-1}(i)P(s^t = s|s^{t-1} = i)$ is joint probability that $\mathbf{X^1}, \mathbf{X^2}, \ldots, \mathbf{X^{t-1}}$ are observed and the state s is reached via state i. Summing this product over all possible states results in all paths reaching s at time t. Since now we know the state distribution at time t, it is easy to compute $\alpha^t(s)$ by incorporating the observed data.

After computing $\alpha^t(s)$, we normalize the probability to compute the conditional state probability distribution

$$P^t(s|\mathbf{X^1}, \mathbf{X^2}, \ldots, \mathbf{X^t}, \lambda) = \frac{\alpha^t(s)}{\alpha^t(1) + \alpha^t(2)} \quad \text{for} \quad s \in \{1, 2\}$$

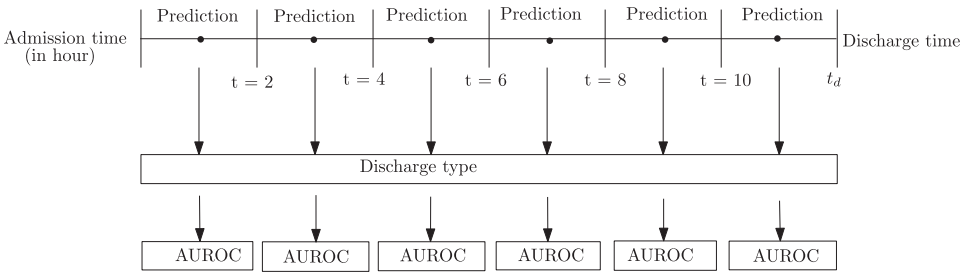

Figure 5. Computing AUROC at each time step.

3.6 Performance evaluation metric

The area under receiver operating characteristic (AUROC) curve was used to compare the performance of temporal and non-temporal methods to model mortality progression using EHR data. The AUROC is a popular performance measure to evaluate the discrimination power of classifiers. Non-temporal models include decision trees, logistic regression, naive Bayes, random forests and support vector machines. In typical scenarios with non-temporal methods, we compute one class probability for an individual patient's visit using aggregated measurements and compare it against the actual outcome to evaluate the AUROC. The HMM computes class probability at each time interval using current and past measurements. Therefore, for comparison, we computed the class probabilities for non-temporal models at each time interval using the measurement recorded in that time interval only. The parameters of non-temporal models were estimated using the data observed in that time series window. Figure 5 illustrates the procedure to compute the AUROC at each time interval. Each model predicts the severity of disease for each time interval, and the predicted risk for each time interval is compared against actual discharge type to compute the AUROC. We employed 10-fold cross-validation to derive the performance measure for both temporal and non-temporal models.

4 Results

4.1 Description of patient characteristics

The descriptive statistics of the selected cohort is shown in Table 1. As expected, the median age of patients with the expired outcome (70 years) is greater than the median age of patients with the non-expired outcome (62 years). The area reflects the location of hospitals. The table also includes minimum, Q1 (first quartile), median, Q3 (third quartile), maximum and mean of each clinical variable stratified by the type of outcome (non-expired and expired). For all three clinical variables, we also performed hypothesis tests to examine the significance of the difference in mean between two outcomes. The results showed that all three variables are significantly important to determine the outcome of the patient hospital stay.

4.2 Performance comparison between proposed temporal framework and non-temporal modeling techniques

Figure 6 shows the AUROC obtained at different time steps prior to discharge time for HMM and non-temporal models on suspected infection data. It is clearly evident from Figure 6 that the AUROC of HMM model is significantly better than non-temporal models. In addition to high performance, HMM model presents a robust increase in the AUROC. The non-temporal models showed high variation in AUROC over time. The primary reason of such behavior is that the non-temporal models estimate the mortality risk by considering only the present measurements of clinical signs but do not incorporate the previous status of the patients. Therefore, the non-temporal models are sensitive to rapid changes in clinical signs.

The difference of AUROC between temporal and non-temporal methods reduces as prediction time moves closer to discharge time. This characteristic indicates that when compared to non-temporal methods, the use of the proposed framework has better prediction advantage in the early stage than the later stage of patient hospital stay. The early identification of a possible bad outcome is an additional feature of the proposed temporal framework.

We performed the trend analysis to clearly inspect the AUROC gap between the proposed temporal framework and the non-temporal methods. Figure 7 includes five plots each comparing an individual non-temporal method

Table 1. Descriptive statistics of the selected cohort (Q1: first Quartile, Q3: third quartile).

Characteristics	Non-expired	Expired
Count, *N* (%)	3629 (93%)	269 (7%)
Gender		
Female, *N* (%)	1667 (93%)	125 (7%)
Age in years, Median	62	70
Area		
Urban, *N* (%)	1442 (93%)	110 (7%)
Rural, *N* (%)	2187 (93%)	159 (7%)
Systolic blood pressure		
Minimum	36	38
Q1	109	96
Median	123	112
Q3	138	132
Maximum	274	225
Mean	123	113
Missing percentage (in 2-h block)	52	51
Glasgow Coma Scale score		
Minimum	2	2
Q1	14	7
Median	15	11
Q3	15	15
Maximum	15	15
Mean	14	11
Missing percentage (in 2-h block)	81	72
Respiratory rate		
Minimum	5	5
Q1	16	17
Median	18	20
Q3	20	24
Maximum	40	40
Mean	19	21
Missing percentage (in 2-h block)	52	43

to the proposed framework. The X-axis represents time prior to discharge and Y-axis shows relative difference of AUROC between the proposed framework and individual methods. In each plot, at the beginning, the AUROC gap is small because the proposed framework takes time to reach steady state from initial state. It is evident from trend lines that the AUROC gap decreases with prediction time moving closer to discharge time. This characteristic emphasizes that the proposed framework is effective for early prediction, which is an important advantage for healthcare applications. The early prediction enables clinicians to inject timely intervention and to efficiently manage hospital resources.

Table 2 summarizes the mean of AUROC obtained by both temporal and non-temporal methods. The mean is calculated by taking the average of all AUROC obtained at different time intervals. The mean AUROC of HMM is 9–12% greater than non-temporal models. The better performance of HMM is attributed to its strength in leveraging longitudinal clinical data.

4.3 Real-time mortality progression: a case study

In this section, we aim to elaborate on another characteristic of the proposed framework that is tracking the mortality progression among suspected infection patients. We developed a mortality progression model by combining the time series data of three clinical variables mentioned in Table 1.

Figure 8 shows the trajectory of three variables (blood pressure, respiratory rate and Glasgow Coma Scale score) along with computed inference (procedure explained in Section 3.5) for the disease severity (or mortality) risk. The inference (red) represents the dynamic behavior of the disease criticality. The region within the dashed rectangle includes the detection point of change in criticality of the patient's condition. This change in state of

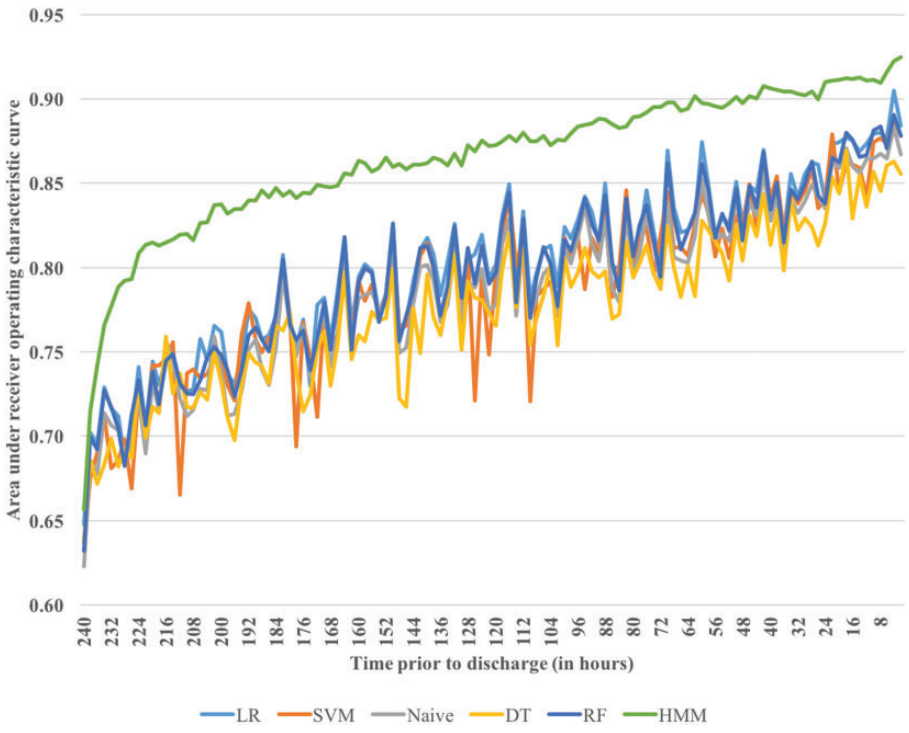

Figure 6. Comparison of area under operating characteristic curve prior to discharge time between HMM and non-temporal methods (LR: logistic regression, SVM: support vector machine, DT: decision tree, RF: random forest, HMM: hidden Markov model).

condition can probably be explained using the drop in blood pressure, continuous low respiration and low Glasgow Coma Scale score. We also noticed in inference that there are further changes in the disease states along the time. The probable reason is the active treatment. The inference using HMM provides critical time information about the change in state and duration of a state. This knowledge can enable practitioners to effectively decide treatment strategies.

For practical usage of clinical tools, the sensitivity and specificity are important measures. Therefore, we computed these performance measures as well. Sensitivity is defined as correctly classifying positive cases, and specificity is correctly classifying negative cases. We examined the performance of our model five days prior to discharge time because it provides enough time for intervention. Figure 9 shows the receiver operating characteristic curve for both temporal and non-temporal models. The sensitivity of our model (0.80) is greater than the sensitivity of all non-temporal models (DT: 0.78, LR: 0.74, Naive: 0.73, RF: 0.74 and SVM: 0.60), which is an important characteristic for diagnosis. But the specificity of our models (0.76) is marginally lower than most of the non-temporal models (DT: 0.78, LR: 0.78, Naive: 0.78, RF: 0.79 and SVM: 0.90). For predictive tools, the sensitivity is an important performance parameter because neglecting the patient who has disease is more detrimental than treating patients that are actually not at high risk. The sensitivity and the specificity of our model, comparing it to the clinically acceptable criteria qSOFA are included in Appendix C, Supplementary Material.

5 Discussion

In this paper, we propose a two state hidden Markov model framework for real-time mortality prediction. We demonstrated the performance of the proposed model on the real-world hospital visit data with suspected infection, obtained from one of the largest EHR repositories in the United States. This paper explains a complete procedure to develop mortality progression from real-world EHR data. The study describes two aspects: the HMM framework to model mortality progression for any disease, and its application to develop a mortality progression model for patients with suspected infection.

This study bridges a significant research gap in that the majority of studies on disease risk models do not leverage longitudinal data of EHR.[3] By using a two state HMM, we combine time series clinical data from EHR and show that the proposed framework performs better than non-temporal models. The modeling foundation of

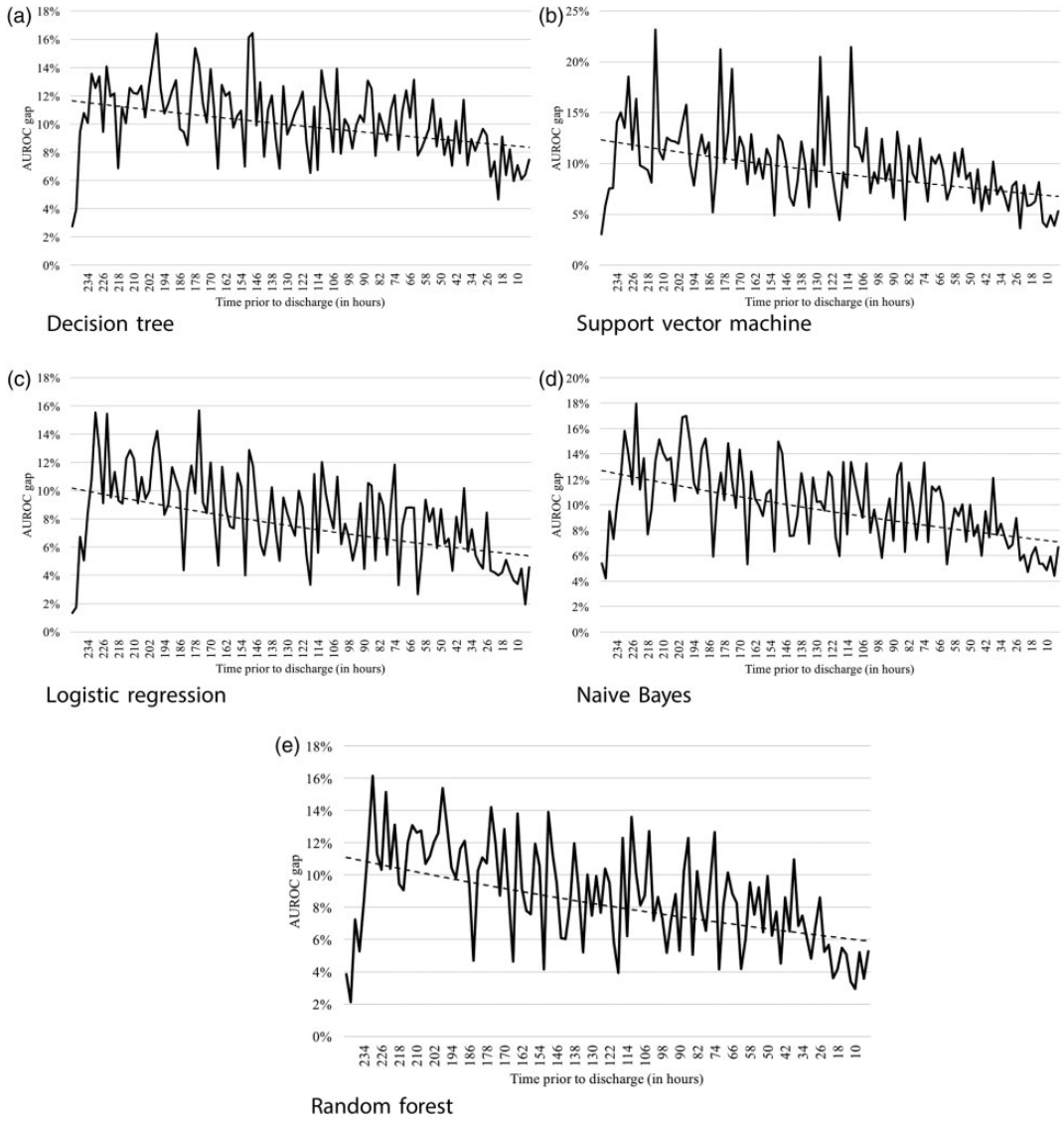

Figure 7. Trend of relative difference in AUROC between proposed framework and non-temporal methods. The dotted line represents the trend line. (a) decision tree; (b) support vector machine; (c) logistic regression; (d) Naive Bayes; (e) random forest.

Table 2. Mean (95% confidence interval (CI)) of AUROC for both temporal and non-temporal models.

Type	Model	Mean AUROC (95% CI)
Non-temporal	Decision tree	0.78 (0.77–0.79)
	Naive Bayes	0.79 (0.78–0.80)
	SVM	0.79 (0.78–0.80)
	Logistic regression	0.80 (0.79–0.80)
	Random forest	0.80 (0.79–0.80)
Temporal	HMM	0.87 (0.86–0.87)

this work is to capture the trend of clinical variables in such a way that is clinically meaningful as well as easy to replicate for aggressively progressing diseases.

Previously, Sukkar et al. developed an HMM-based disease risk progression model for slow progressing disease.[27] The author employed six-state HMM to model Alzheimer's disease. However, the multi-state model

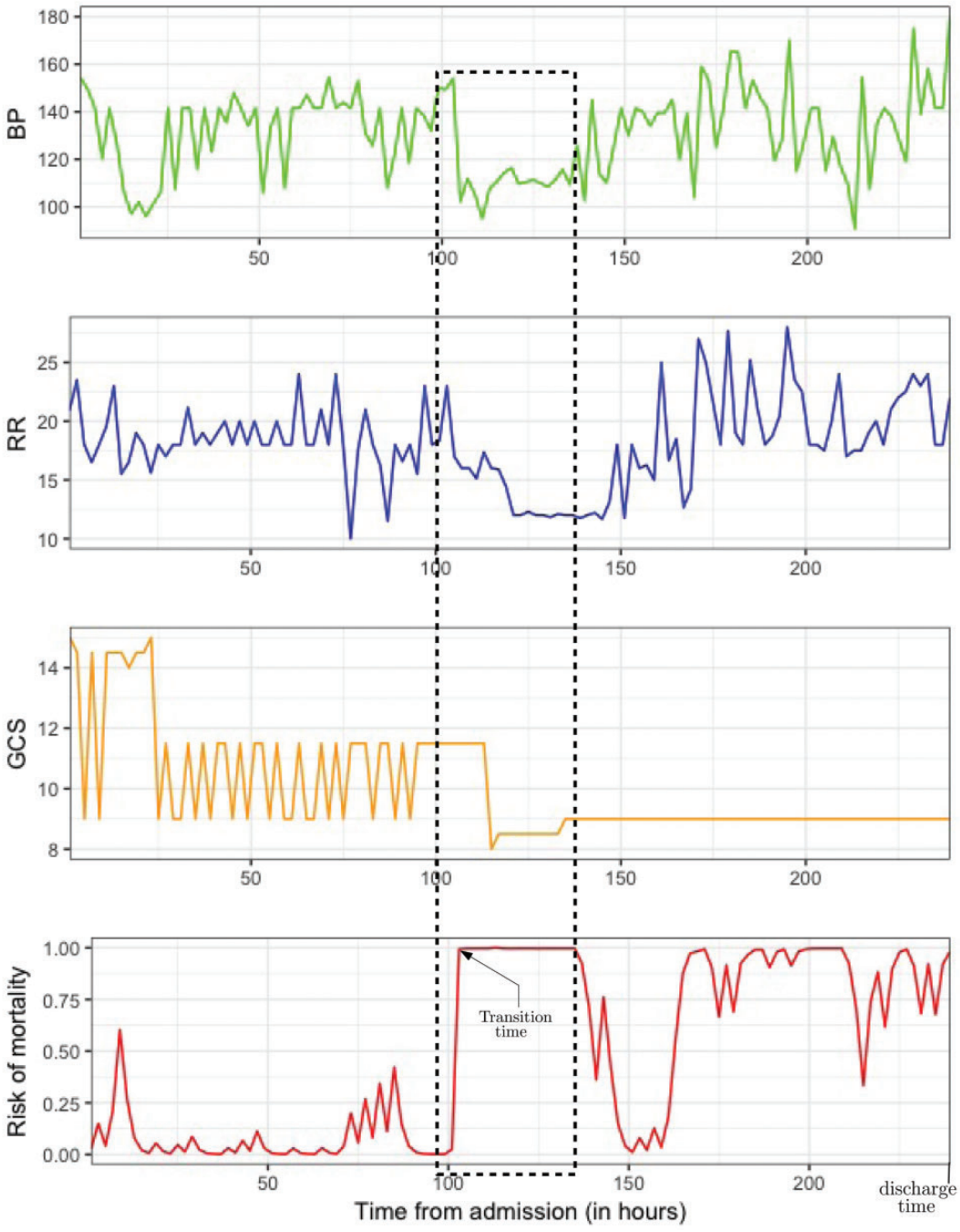

Figure 8. Example encounter features and mortality trajectory. BP: systolic blood pressure, RR: respiratory rate, and GCS: Glasgow Coma Scale score.

possesses its own challenges such as learning complexity. This learning problem becomes significant especially for EHR where high volumes of missing data are present. This study describes a step-by-step process to build a time series model from EHR data. This study directs future clinical researchers to consider time variations in variables instead of merely using aggregated values (mean/median). The proposed approach also provides time-related insights that enable practitioners to understand the trajectory of disease.

The mortality prediction model designed using traditional non-temporal approaches (decision tree, logistic regression, naive Bayes, random forest and support vector machine) is very sensitive to short-time intervention effects. The prediction in non-temporal models is computed using the latest measurements. Therefore, temporary

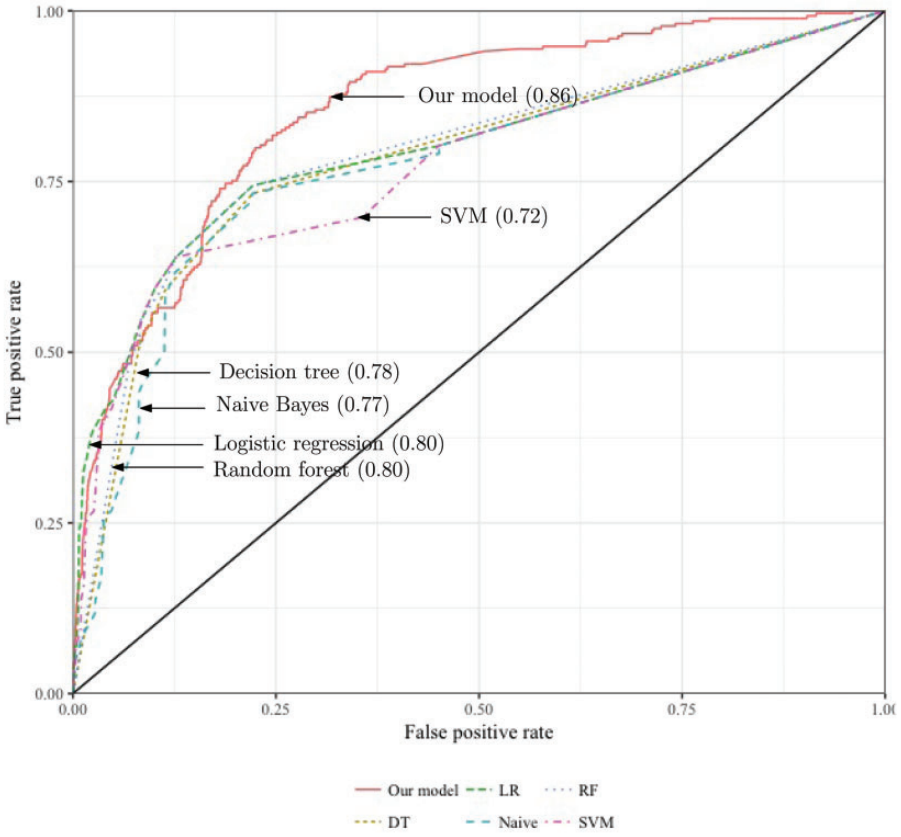

Figure 9. Receiver operating characteristic curve. Prediction time is five days prior to discharge. SVM: Support vector machine.

change in clinical signs due to intervention can significantly affect the prediction of mortality risk. Such predictions could be misleading and can lead to poor treatment decisions.[31] For example, vasopressin is used to treat septic patients. The intake of vasopressin can induce temporary fall in heart rate. Hence, deciding treatment strategy solely based on instant changes prompted by interventions can lead to poor treatment decisions. Therefore, instead of relying only on current clinical observations, the more robust mortality risk could be obtained by incorporating the temporal behavior of the clinical signs. Consequently, the change in mortality risk does not depend on instant clinical signs; rather, it depends on the continuous trend of clinical signs. Using the proposed temporal framework, the more robust mortality prediction can be achieved by combining the trends of individual clinical signs.

The proposed framework provides continuous update on the health of a patient. The status of the patient's health gets updated periodically with a newly observed set of clinical observations. This real time information may facilitate early detection of acute events, discharge planning and managing limited resources of hospitals. Although the case study data in this paper only includes vital signs due to their dense availability in EHRs, the lab results could also be incorporated. The periodic update on the mortality risk requires knowledge of all clinical signs at each time period. In case any clinical variable is not available, the possible alternative is to use the last available measurement to infer about mortality risk.

The mortality progression model, developed as a part of this study, showed a balance for both sensitivity and specificity. The developed model provides an alternative to qSOFA, a clinically accepted criteria. Our model only uses three easy-to-measure non-invasive clinical measurements to infer about mortality progression. The proposed model also helps understand *time* at which patient criticality changes and *time duration* for which patient was in a specific state. This model equips physicians with early diagnosis in order to provide timely treatment.

This study has a few limitations. We assumed that HMM parameters are stationary or do not vary over time. This is a legitimate assumption for fast progressing diseases. Another assumption of this study is that hidden states follow Markovian property. The Markovian property implies that given the previous state, the current state is independent of other states. The intuition behind that assumption is that the previous state provides a summary

of what had happened one step prior to current time. Our model limits the range of input values (SBP: 35-274, RR: 5-40, and GCS: 2-15). Although the clinical values beyond this range are less likely, they may occur. The advantage of the proposed model comes at the expense of increased computational complexity compared to qSOFA. The latter is much simpler. However, by using a simple computer code, we can automate the mortality risk estimation process.

Acknowledgements

This work was conducted with data from the Cerner Corporation's Health Facts database of electronic medical records provided by the Oklahoma State University (OSU) Center for Health Systems Innovation (CHSI). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Cerner Corporation. The authors also like to acknowledge the data extraction support from Elvena Fong and timely suggestions from Zhuqi Miao, Ph.D. both from OSU-CHSI. We also want to acknowledge the time and constructive feedback of unknown referees to improve the reproducibility and organization of the work.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Akash Gupta https://orcid.org/0000-0002-1306-0390

Supplemental material

Supplemental material for this article is available online.

References

1. Matheny ME, Miller RA, Ikizler TA, et al. Development of inpatient risk stratification models of acute kidney injury for use in electronic health records. *Med Decis Mak* 2010; **30**: 639–650.
2. Tabak YP, Sun X, Nunez CM, et al. Using electronic health record data to develop inpatient mortality predictive model: acute laboratory risk of mortality score (alarms). *J Am Med Inform Assoc* 2013; **21**: 455–463.
3. Goldstein BA, Navar AM, Pencina MJ, et al. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc* 2017; **24**: 198–208.
4. Charles D, Gabriel M and Furukawa MF. Adoption of electronic health record systems among us non-federal acute care hospitals: 2008–2012. *ONC Data Brief* 2013; **9**: 1–9.
5. Singer M, Deutschman CS, Seymour CW, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *J Am Med Inform Assoc* 2016; **315**: 801–810.
6. Freund Y, Lemachatti N, Krastinova E, et al. Prognostic accuracy of sepsis-3 criteria for in-hospital mortality among patients with suspected infection presenting to the emergency department. *J Am Med Inform Assoc* 2017; **317**: 301–308.
7. Fleischmann C, Thomas-Rueddel DO, Hartmann M, et al. Hospital incidence and mortality rates of sepsis: an analysis of hospital episode (drg) statistics in Germany from 2007 to 2013. *Deutsches Ärzteblatt Int* 2016; **113**: 159.
8. Torio CM and Andrews RM. National inpatient hospital costs: the most expensive conditions by payer, 2011: statistical brief# 160, 2006.
9. Seymour CW, Liu VX, Iwashyna TJ, et al. Assessment of clinical criteria for sepsis: for the third international consensus definitions for sepsis and septic shock (sepsis-3). *J Am Med Inform Assoc* 2016; **315**: 762–774.
10. Saltzman JR, Tabak YP, Hyett BH, et al. A simple risk score accurately predicts in-hospital mortality, length of stay, and cost in acute upper GI bleeding. *Gastrointest Endosc* 2011; **74**: 1215–1224.
11. Ramchandran KJ, Shega JW, Von Roenn J, et al. A predictive model to identify hospitalized cancer patients at risk for 30-day mortality based on admission criteria via the electronic medical record. *Cancer* 2013; **119**: 2074–2080.
12. Bone RC, Balk RA, Cerra FB, et al. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Chest* 1992; **101**: 1644–1655.
13. Subbe C, Kruger M, Rutherford P, et al. Validation of a modified early warning score in medical admissions. *QJM* 2001; **94**: 521–526.
14. Vincent JL, Moreno R, Takala J, et al. The sofa (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. *Intens Care Med* 1996; **22**: 707–710.

15. Gupta A, Liu T, Shepherd S, et al. Using statistical and machine learning methods to evaluate the prognostic accuracy of sirs and qsofa. *Healthcare Informat Res* 2018; **24**: 139–147.
16. Calvert J, Mao Q, Hoffman JL, et al. Using electronic health record collected clinical variables to predict medical intensive care unit mortality. *Annals Med Surg* 2016; **11**: 52–57.
17. Henry KE, Hager DN, Pronovost PJ, et al. A targeted real-time early warning score (trewscore) for septic shock. *Sci Translat Med* 2015; **7**: 299ra122–299ra122.
18. van Wyk F, Khojandi A, Mohammed A, et al. A minimal set of physiomarkers in continuous high frequency data streams predict adult sepsis onset earlier. *Int J Med Informat* 2019; **122**: 55–62.
19. Peelen L, de Keizer NF, de Jonge E, et al. Using hierarchical dynamic Bayesian networks to investigate dynamics of organ failure in patients in the intensive care unit. *J Biomed Inform* 2010; **43**: 273–286.
20. Cai X, Perez-Concha O, Coiera E, et al. Real-time prediction of mortality, readmission, and length of stay using electronic health record data. *J Am Med Inform Assoc* 2015; **23**: 553–561.
21. Shickel B, Tighe PJ, Bihorac A, et al. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J Biomed Health Inform* 2017; **22**: 1589–1604.
22. Ching T, Himmelstein DS, Beaulieu-Jones BK, et al. Opportunities and obstacles for deep learning in biology and medicine. *J Royal Soc Interface* 2018; **15**: 20170387.
23. Raghu A, Komorowski M, Celi LA, et al. Continuous state-space models for optimal sepsis treatment-a deep reinforcement learning approach. *arXiv preprint arXiv:170508422* 2017.
24. Kam HJ and Kim HY. Learning representations for the early detection of sepsis with deep neural networks. *Comput Biol Med* 2017; **89**: 248–255.
25. Liu YY, Ishikawa H, Chen M, et al. Longitudinal modeling of glaucoma progression using 2-dimensional continuous-time hidden Markov model. In: *International conference on medical image computing and computer-assisted intervention*, September 2019. Berlin, Heidelberg, pp.444–451.
26. Vairavan S, Eshelman L, Haider S, et al. Prediction of mortality in an intensive care unit using logistic regression and a hidden Markov model. In: Murray A (ed) *2012 Computing in cardiology*. Krakow, Poland: IEEE, pp.393–396.
27. Sukkar R, Katz E, Zhang Y, et al. Disease progression modeling using hidden markov models. In: *2012 Annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, August 2012. San Diego: IEEE, pp.2845–2848.
28. McGhee TL, Weaver P, Solo S, et al. Vital signs reassessment frequency recommendation. *Nursing Manage* 2016; **47**: 11–12.
29. Gupta A, Liu T and Shepherd S. Clinical decision support system to assess the risk of sepsis using tree augmented Bayesian networks and electronic medical record data. *Health Informat J* 2019: 1460458219852872.
30. Rabiner LR. A tutorial on hidden markov models and selected applications in speech recognition. *Proc IEEE* 1989; **77**: 257–286.
31. Windt J, Ardern CL, Gabbett TJ, et al. Getting the most out of intensive longitudinal data: a methodological review of workload–injury studies. *BMJ Open* 2018; **8**: e022626.