DyCRS: Dynamic Interpretable Postoperative Complication Risk Scoring

Wen Wang
Carnegie Mellon University
Pittsburgh, PA, USA
wenw3@andrew.cmu.edu

Han Zhao
Carnegie Mellon University
Pittsburgh, PA, USA
han.zhao@cs.cmu.edu

Honglei Zhuang
University of Illinois at
Urbana-Champaign
Champaign, IL, USA
hzhuang3@illinois.edu

Nirav Shah
NorthShore University HealthSystem,
University of Chicago Pritzker School
of Medicine
Chicago, IL, USA
NShah2@northshore.org

Rema Padman
Carnegie Mellon University
Pittsburgh, PA, USA
rpadman@cmu.edu

ABSTRACT

Early identification of patients at risk for postoperative complications can facilitate timely workups and treatments and improve health outcomes. Currently, a widely-used surgical risk calculator online web system developed by the American College of Surgeons (ACS) uses patients' *static* features, e.g. gender, age, to assess the risk of postoperative complications. However, the most crucial signals that reflect the actual postoperative physical conditions of patients are usually real-time *dynamic* signals, including the vital signs of patients (e.g., heart rate, blood pressure) collected from postoperative monitoring. In this paper, we develop a dynamic postoperative complication risk scoring framework (DyCRS) to detect the "at-risk" patients in a real-time way based on postoperative sequential vital signs and static features. DyCRS is based on adaptations of the Hidden Markov Model (HMM) that captures hidden states as well as observable states to generate a real-time, probabilistic, complication risk score. Evaluating our model using electronic health record (EHR) on elective Colectomy surgery from a major health system, we show that DyCRS significantly outperforms the state-of-the-art ACS calculator and real-time predictors with 50.16% area under precision-recall curve (AUCPRC) gain on average in terms of detection effectiveness. In terms of earliness, our DyCRS can predict 15hrs55mins earlier on average than clinician's diagnosis with the recall of 60% and precision of 55%. Furthermore, Our DyCRS can extract interpretable patients' stages, which are consistent with previous medical postoperative complication studies. We believe that our contributions demonstrate significant promise for developing a more accurate, robust and interpretable postoperative complication risk scoring system, which can benefit more than 50 million annual surgeries in the US by substantially lowering adverse events and healthcare costs.

CCS CONCEPTS

• **Applied computing** → **Health informatics**; • **Information systems** → **Data mining**; • **Computing methodologies** → *Machine learning*.

KEYWORDS

Real-time risk score; Postoperative complications; Hidden Markov Model; Interpretability

1 INTRODUCTION

Over 50 million surgeries are performed in the U.S. annually, with surgical hospitalization costs rising to more than 30% of all inpatient costs for the U.S. healthcare system (National Health Statistics Reports, 2010). Complication rates range between 6%-47% and average cost per complication ranges from $675 to $2,800 for a urinary tract infection (UTI), $27,000 for a surgical site infection (SSI) and over $50,000 for ventilator-associated pneumonia [22, 29, 35]. Thus, postoperative complications are a critical driver of healthcare expenditures [3, 13]. Moreover, postoperative patients who are worked up without complications unnecessarily incur significant morbidity and healthcare expenditures [10]. Early identification of patients at risk for postoperative complications, leading to timely and accurate workups, may significantly reduce adverse events and readmission rates [23, 41].

The current state-of-the-art postoperative complications risk calculator is the American College of Surgeons National Surgical Quality Improvement Program model (ACS NSQIP model [5, 8]). ACS NSQIP is a widely used online web system as shown in Fig. 1. It is a national, validated, outcome-based, risk-adjusted, and peer-controlled program for measurement, benchmarking and enhancement of the quality of surgical care [8]. It is designed for pre-operative assessment which is necessary prior to the majority of

Figure 1: Preoperative Risk Assessment ACS NSQIP Surgical Risk Calculator Online Web System.

elective surgical procedures, to discuss surgery risk with patients and ensure that the patient is fit to undergo surgery [33]. It uses patients' static features (e.g. gender, etc.) to calculate the risk of postoperative complications.

On the other hand, real-time postoperative assessment is also very important. Postoperative patients could be in a very fragile state and vulnerable to postoperative complications due to surgical trauma [17]. Although the patients' vital signs are closely monitored after surgery, it heavily relies on the clinicians to discover the potential risk of further complications from the dynamics of collected data. If the sign of complications can't be discovered in time, the health condition will deteriorate rapidly and even lead to death [37]. Thus, automatically measuring the risk of complications in a real-time way is necessary to improve early diagnosis and avoid related adverse events. However, the ACS NSQIP model does not work in postoperative assessment, since it lacks dynamic and real-time capabilities to capture postoperative clinical sequence transition dependency, as it is a static and one-shot prediction model. Better methods are needed to detect the "at-risk" patients in a real-time way for timely postoperative assessment.

Fig. 2 shows an example system for real-time postoperative complication risk assessment. In common postoperative care, once the surgery is done, multiple vital signs (e.g. blood pressure, white blood cell counts, etc.) of the patients are taken sequentially and closely. Besides, for any elective surgery, patients on admission will provide their basic information (e.g. gender, comorbidities). We aim to construct a real-time complication risk score based on these sequential vital signs and patients' static features as well.

However, there are several challenges in constructing a real-time risk detection system in the postoperative setting:

- First, patients' complication risk varies over time, but it depends on all of the previous vital signs since there is a strong sequential dependency among vital signs. Therefore widely used real-time prediction methods (e.g. sliding window methods [20, 28, 40]) are limited in measuring the long-term dependency and can only capture patients' health status changes within a fixed duration. So some meaningful and
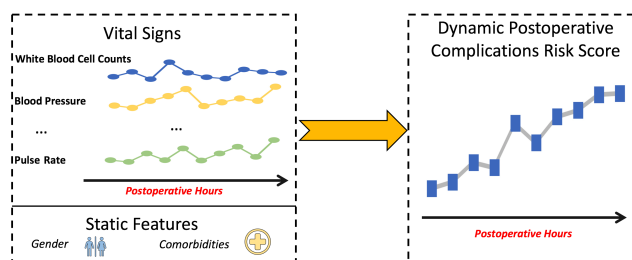

Figure 2: Illustration of the Input and Output of Postoperative Risk Assessment Real-time System.

representative clinical states that may be informative to the clinically worsening can be discarded, therefore may fail.
- Second, medical data, especially surgery data, is very sparse due to the privacy issue [30]. Besides, it is costly for hospitals to maintain surgery electronic health warehouses since a large amount of well-trained staff is needed to take care of the monitoring and recording [15]. Thus, it is very hard to train an extremely complicated model such as a deep neural network-based model from a limited amount of training data.
- More importantly, the model must be interpretable for clinicians [11]. It should also be consistent with the current results of medical research. Only in this way, the clinicians can trust the model [14].

Our contributions. In this paper, we propose a **dy**namic interpretable postoperative **c**omplication **r**isk **s**coring (DyCRS) framework to detect "at-risk" patients in real-time. DyCRS is able to generate a real-time, probabilistic, complication risk score based on the patients' postoperative vital signs and static features (e.g. gender). DyCRS is based on adaptations of Hidden Markov Model (HMM). Specifically, we have the following four contributions:

- From the application perspective, different from the preoperative assessment–ACS NSQIP risk calculator online web system, we propose a dynamic postoperative risk assessment framework DyCRS to generate real-time risk score. The score given by our system can be sent to clinicians' mobile phones in the future. It is convenient for clinicians to monitoring the risk of patients at anytime and anywhere.
- From the methodology perspective, DyCRS combines two existing adaptations of the Hidden Markov Model (HMM) together to fit our setting: capturing observable states in addition to hidden states [2] in the state space, and incorporating static features in addition to sequential features [32]. Based on that, we further develop a pipeline to convert the discrete hidden states to the continuous risk score.
- We evaluate the effectiveness of DyCRS in comparison with 1. machine learning and deep learning real-time classifiers, 2. the widely used preoperative assessment – ACS NSQIP surgical risk calculator, 3. the state-of-care postoperative assessment commonly used by clinicians in hospitals–fever criterion. Among all the competitors, the proposed DyCRS significantly outperforms these baselines in terms of effectiveness, earliness, and clinical significance.

- In contrast to existing methods that are almost "black-box" predictors, we further demonstrate that our model can extract *interpretable* patients' stages, which are consistent with previous medical postoperative complication studies.

2 PROBLEM SETUP

Our task is predicting complication risk in a real-time way based on postoperative patients' sequential and static features. We first describe the data and then task formulation.

2.1 Dataset

The NorthShore University HealthSystem maintains an electronic data warehouse (EDW) that captures information entered into patients' electronic health records. Complication data were extracted on all patients who underwent elective Colectomy surgery between January 1, 2007, and December 31, 2013.

Features. Our complication data include three types of features.

- **Sequential features**. Five common postoperative vital signs are taken sequentially and closely to monitor the postoperative health condition during the inpatient stay: *white blood cell counts* (1000s per microliter of blood), *systolic blood pressure* (mm Hg), *diastolic blood pressure* (mm Hg), *pulse rate* (pulses per minute) and *body temperature* ($^\circ F$).
- **Static features**. Patients' static features are also available in EHR. Suggested by clinicians, we incorporate two informative static features for Colectomy surgery including gender and uncomplicated diabetes in our model. 38 static features in total are available for the comparison experiments.
- **Time-to-events**. During the inpatient stay, the time to clinical worsening(first occurrence of the postoperative complications) is available for some patients. For clinically stable patients, the time to discharge from the hospital is also available. We have two events: clinical worsening or discharge from the hospital.

Data preprocessing. We first preprocess five vital signs to occur at synchronous four-hour time intervals. Suggested by clinicians, we use the maximum value of vital signs within each 4-hour interval, following common practice for tracking the onset of complications in the hospital. Second, we use a linear interpolation method to deal with missing values for each patient's vital signs because we assume the patients' health condition will not change suddenly. Finally, our dataset includes 526 patients with 12,666 time step data points, each time step is associated with 5 vital signs. 119 (23%) patients develop postoperative complications after surgery.

Preliminary analysis. Before we start to build a model, we first do a simple data analysis to demonstrate the importance of developing a real-time postoperative assessment system. We select a representative static feature utilized by the ACS NSQIP risk calculator system (i.e. age) and plot a histogram with respect to the postoperative complications rates of patients. In comparison, we also select a representative dynamic feature (i.e. average 24-hour postoperative temperature before events) obtained from patients' real-time vital signals to plot a histogram with respect to postoperative complications rates. As shown in Fig. 3, the correlation between static features of patients and their postoperative complication risk is insignificant, whereas the real-time vital signals present a strong

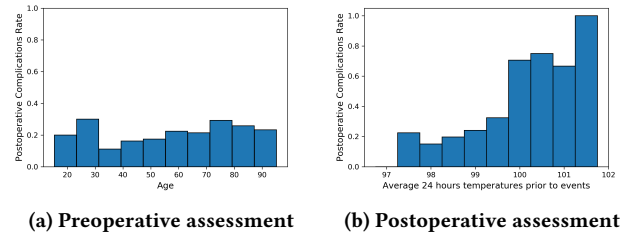(a) Preoperative assessment   (b) Postoperative assessment

Figure 3: Preliminary data analysis of preoperative and postoperative assessment. This shows the strong positive association between postoperative information and complications in contrast to no obvious association for preoperative information and complications.

positive correlation. The comparison again verifies that the traditional risk assessment tool relying on static features may not have sufficient information to provide an accurate prediction of postoperative complications. However, with postoperative real-time vital signals available, which are more able to reflect the actual physical condition of the postoperative patients, it is possible to make a more informative prediction of possible complications and provide more meaningful support for life-saving clinical decision making.

2.2 Task Formulation

In this section, we will first describe our modeling intuition, and then explain the adaptations of our model based on HMM.

Modeling intuition. As shown in Fig 4, we randomly pick up a clinical worsening and a discharged patient and plot their postoperative vital signs and time-to-event. Once the surgery is done, the patient will be moved to the ward and enter into the monitoring period. In the monitoring period, vital signs are taken sequentially and closely to see if the patient shows any signs of postoperative complications. Therefore, in this period, the patient bears some *latent* risk of developing complications, but the event does not occur actually. To some point, the patient is clinical worsening and complication occurs, the patient is transferring from bearing some *latent* risk to *observed* adverse event with the certainly highest risk. On the other hand, clinically stable patient will bear *latent* risk of developing complications in the monitoring period, to some point, the clinicians remove the patients' *latent* risk and announce that the patient can be discharged from the hospital, therefore the patient transferring from bearing some *latent* risk to *observed* discharge event with certainly no risk.

Therefore, the patients' health condition can be represented by the risk stages w.r.t postoperative complications during the inpatient stay. In the monitoring period, the patients bear some *latent* risk, which is an intermediate level of risk (e.g. low, medium, high risk w.r.t complications). The *observed* clinical worsening event is the highest risk stage since the patient has already encountered an adverse event. The *observed* discharge event is no risk stage since the risk is removed by clinicians. Thus, with the increasing of severity, patients risk stages can be summarized as: no risk (*observed* discharge event), intermediate level risk (*latent* risk stages), highest risk (*observed* clinical worsening event) (See the conceptual structure of modeling in Fig 5). No risk and highest risk stages are
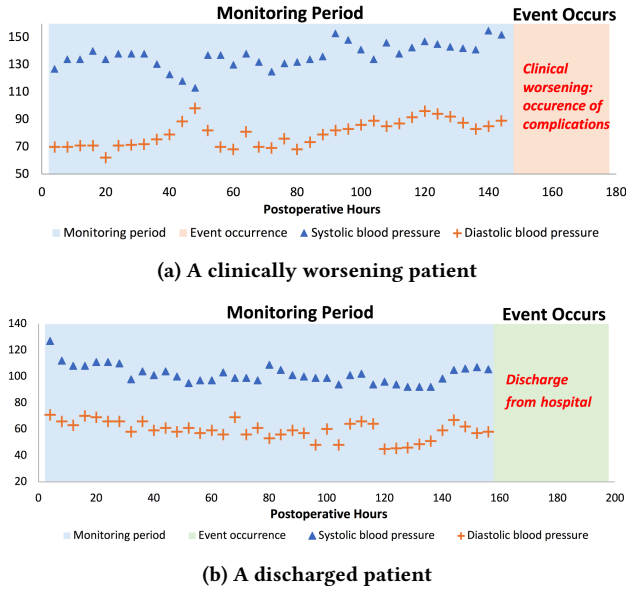
(a) A clinically worsening patient

(b) A discharged patient

Figure 4: Visualization of two patients' vital signs and time-to-event to demonstrate the modeling intuition.

observed, while the intermediate-risk stages are latent. Only the vital signs can be observed to infer the latent risk stages. So basically, we need a model to capture the transition relationship between all risk stages including latent and observed, and the relationship between the multivariate vital signs and latent risk stages.

Besides, our goal is to propose a real-time risk score for patients. After we learn the above two relationships, given a new patients' vital signs, we can first decode the probability distribution of *latent* risk stages, and using the transition relationship between *latent* risk stage and *observed* clinically worsening event to construct a risk score that represents the probability the patients will finally encounter the worsening stage rather than discharge stage.

DyCRS. Hidden Markov Model (HMM) ([34]) is a well-known probabilistic graphical model to capture the relationship between *latent* states sequences and *observed* observations sequences. The latent states will be represented by a multivariate Gaussian distribution to link with vital signs, this distribution is known as emission probability. Besides, the probabilistic nature of HMM also provides us with better interpretability compared to the black-box machine learning algorithm [6], this is extremely important in health care decision making contexts. Thus, it is very suitable to use HMM in our case. We combine two existing adaptations of HMM and proposed new adaptation together to fit in our setting:

- In our task, two observed states are available for each patient: *clinical worsening event* or *discharge event*. Therefore, the patient's postoperative evolving risk stages trajectory will be hidden until it ends at either *clinical worsening state* or *discharge state*. Thus, we adopt the adapted HMM to capture observed terminating states [2].
- In addition to postoperative sequential features, the patient's static features are also available and informative for postoperative complications risk assessment. Hence, we also combine

the adapted HMM to incorporate patient static features in emission probability [32].

- Besides the above two existing adaptations, we also develop a pipeline to convert hidden states to the real-time risk score. Given the estimated parameters of HMM(e.g. transition matrix, emission probability), for a new patient vital signs, the traditional decoding problem of HMM can only tell us the most likely hidden states. However, the hidden states are discrete(e.g. low risk state) instead of a real number, thus it lacks the accuracy to quantify the complication risk. Besides, it is also meaningless and abstract to show the patient is in low risk stage, a real number that tells us what is the probability the patients will get to complication event will be more meaningful and straightforward.
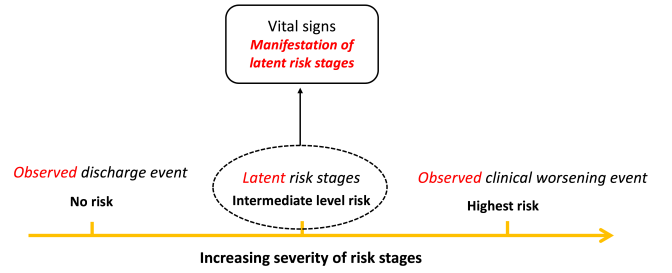

Figure 5: The conceptual relationship between all states including latent and observed states, and the relationship between latent states and vital signs.

3 MODEL AND ALGORITHM

In this section, we define our DyCRS model formally. At a high level, our model contains two phases: (*i*) a learning phase to learn the transition relationship between hidden states and observed states, and the relationship between hidden states and observations (e.g. vital signs); (*ii*) a subsequent inference phase to convert the hidden state to risk score.

3.1 Model Description

3.1.1 State Space. Given the three types of patient information, we construct a discrete HMM $\mathcal{H}$ with hidden states and observed states as shown in Fig. 6. Since three hidden states have a natural medical explanation (e.g. low risk, medium risk, high risk), we choose three hidden states in our model, but it can easily generalize to any number of hidden states. In summary, the state space of $\mathcal{H}$ contains five discrete states: discharge event (D), three latent states that correspond to increasing severity of patients' health status (low risk (L), medium risk (M), high risk (H)), and clinical worsening event: occurrence of postoperative complications (C). Note that in our model, both C and D are absorbing states and they can be observed, whereas the other three states (L, M and H) are transient and hidden. An absorbing state is a state that, once entered, cannot be left. Therefore, we change the common Markov chain to the absorbing Markov chain to incorporate the observed terminating states. The absorbing Markov process will start at one of the three transient states and has to finally enter either state C or D. We use $\mathcal{S}$ to denote the set of states in $\mathcal{H}$.
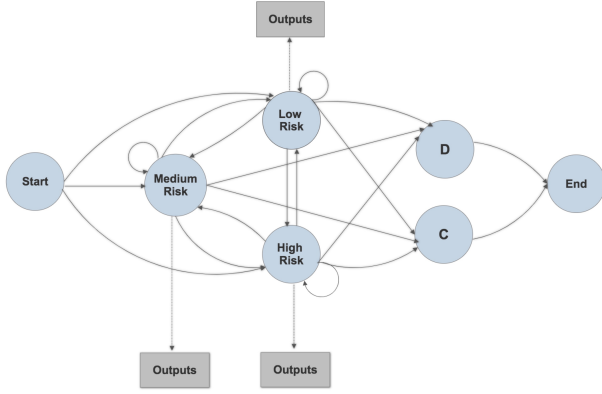
Figure 6: The probabilistic finite state machine of the HMM $\mathcal{H}$ used in our DyCRS system. Low, medium and high risk states are hidden states that represent the intermediate level risk. C represents the clinical worsening event, and D represents the discharge event. C is the highest risk state, D is no risk state, both are observed.

3.1.2 Observation Space. Observable outcomes are defined by patients' five postoperative vital signs: *white blood cell counts* (1000s per microliter of blood), *systolic blood pressure* (mm Hg), *diastolic blood pressure* (mm Hg), *pulse rate* (pulses per minute) and *body temperature* ($°F$). As a result, each patient's observation at a specific time step can be described by a 5-dimension continuous vector $o \in \mathbb{R}^d$ with $d = 5$.

3.1.3 Transition Probability Matrix. For an HMM $\mathcal{H}$ with m states including hidden and observed states, the transition matrix $T \in \mathbb{R}_+^{m \times m}$ characterizes the probability of the model to transit from one state to another. Formally, we define

$$T_{ij} := \Pr(s' = i \mid s = j), \quad \forall i, j \in \mathcal{S}. \tag{1}$$

By definition, it is easy to verify that T is a column-stochastic matrix where each column sums to 1, i.e., $\mathbf{1}_m^T T = \mathbf{1}_m^T$.

As shown in Fig. 6, our HMM $\mathcal{H}$ has 5 states, of which two are absorbing states. For the convenience of discussion, we fix the order of states to be $\mathcal{S} = \{D, L, M, H, C\}$. Under this order, to represent that both the C and D are absorbing states, the transition matrix T is constrained to have the following structures:

$$T_{11} = T_{55} = 1, \quad T_{i1} = 0, \ \forall i \neq 1, \quad T_{j5} = 0, \ \forall j \neq 5. \tag{2}$$

The remaining elements of T will be learned from the sequences of observations and shared across all the patients.

3.1.4 Emission Probability Model. In this section, we describe in detail the continuous emission probability model in our HMM $\mathcal{H}$, which takes into account both vital signs and static features. To simplify the notation, in the following derivation we omit the time index t and the patient index when they are clear from the context. Let $s \in \{0, 1\}^m$ be a one-hot vector that encodes the index of a state, and $r \in \mathbb{R}^p$ be a patient static features vector that describes patient-specific information, e.g., gender, uncomplicated diabetes, etc. We assume that the observation follows a parametrized multivariate Gaussian distribution whose mean parameter is a linear function

of both the hidden state s and the patient heterogeneity vector r:

$$p(o \mid s) = \mathcal{N}(Ws + Mr, \sigma^2 I_d), \tag{3}$$

where $W \in \mathbb{R}^{d \times m}$ is the weight matrix of the state-observation relationship and $M \in \mathbb{R}^{d \times p}$ is the static features-observation matrix that characterizes how patient's heterogeneous attribute affects the observation. Without loss of generality, we assume an isotropic covariance matrix in the above Gaussian distribution where the variance of each dimension is given by σ^2. This assumption is by no means restrictive and can be easily relaxed.

Remark. Note that in the Gaussian model we define in (3), the variance parameter σ^2 is fixed apriori and in the learning process our goal is to infer both weight matrices W and M, where s is a one-hot vector. In this case, each column vector of the weight matrix W has a natural interpretation: it represents the mean value of the observation vector given by the specific hidden state.

3.1.5 Initial Distribution. We learn the initial distribution of $\mathcal{H}$ during the learning phase. Since both D and C are absorbing states, their corresponding initial probabilities are fixed to be 0. The other three initial probabilities corresponding to L, M and H will be estimated by the EM algorithm.

3.2 DyCRS Learning

Given a set of patients, where each patient corresponds to an observation sequence $\mathbf{o}_i$, we first learn the model parameters T, W and M that maximize the probability of $\mathcal{H}$ generating the data. Define $\theta := \{T, W, M\}$ to be the set of model parameters. We use the maximum likelihood principle (MLE) to formalize the learning problem as an optimization problem:

$$\max_{\theta} \quad \sum_{i=1}^{n} \log p(\mathbf{o}_i \mid \theta), \qquad \text{subject to} \quad T \geq 0, \ \mathbf{1}_m^T T = \mathbf{1}_m^T \tag{4}$$

where $T \geq 0$ is meant to be elementwise. Due to the existence of hidden states, it is not hard to see that the optimization problem (4) is nonconvex. In light of this intrinsic hardness, we propose a EM algorithm for learning the model parameters.

3.2.1 Learning Transition Probability Matrix and Initial Distribution. To clearly state the algorithm for learning the transition probability matrix, given an observation sequence $\mathbf{o}$ of length t, we first define what we call the forward and backward probability.

Definition 3.1 (Forward probability). Given an HMM $\mathcal{H}$ and a sequence $\mathbf{o}$ of observations of length t, $\forall j \in [t]$, we define the *forward probability* at time step j as $\alpha_j(\mathbf{o}) := p(s_j \mid o_{1:j}) \in \mathbb{R}_+^m$.

Definition 3.2 (Backward probability). Given an HMM $\mathcal{H}$ and a sequence $\mathbf{o}$ of observations of length t, $\forall j \in [t]$, we define the *backward probability* at time step j as $\beta_j(\mathbf{o}) := p(o_{j+1:t} \mid s_j) \in \mathbb{R}_+^m$.

For a specific observation vector $o \in \mathbb{R}^d$, we define the observable operator matrix A_o w.r.t. o as follows:

$$A_o := \text{diag}(p(o \mid s = 1), \dots, p(o \mid s = m)) \cdot T \in \mathbb{R}_+^{m \times m}, \tag{5}$$

where the probability density function $p(\cdot \mid \cdot)$ is given by (3). Furthermore, we also need to compute the posterior belief state at each time step j, i.e., $\gamma_j := p(s_j \mid \mathbf{o})$. With the forward and backward probabilities, the belief state can be easily computed as follows:

$$\gamma_j(\mathbf{o}) = p(s_j \mid o_{1:t}) \propto p(s_j \mid o_{1:j}) \cdot p(o_{j+1:t} \mid s_j) = \alpha_j(\mathbf{o}) \circ \beta_j(\mathbf{o}), \tag{6}$$

where $\circ$ denotes the Hadamard product, i.e., elementwise product.

Similarly, we can also compute the probability of state transition in the following way:

$$p(s_j, s_{j+1} \mid \mathbf{o}) \propto p(s_j \mid o_{1:j}) \cdot p(o_{j+1} \mid s_{j+1}) \cdot p(s_{j+1} \mid s_j) \cdot p(o_{j+2:t} \mid s_{j+1})$$

$$= A_{o_{j+1}} \circ \left(\alpha_j(\mathbf{o}) \beta_{j+1}^T(\mathbf{o}) \right). \qquad (7)$$

To estimate the initial distribution, we use the averaged belief state over all the sequences:

$$\hat{\pi} = \frac{1}{n} \sum_{i=1}^{n} \gamma_1(\mathbf{o}_i). \qquad (8)$$

To estimate the transition matrix, we sum all the expected state transition counts over all the time steps in all the observation sequences:

$$\hat{T} \propto \sum_{i=1}^{n} \sum_{j=1}^{t_i-1} p(s_j, s_{j+1} \mid \mathbf{o}_i). \qquad (9)$$

3.2.2 Learning Emission Probability Model. To learn the weight matrices W and M in the emission model, we maximize the expected log-likelihood function, which can be simplified to the following weighted least square minimization problem:

$$\min_{W, M} \qquad \sum_{i=1}^{n} \sum_{j=1}^{t_i} \sum_{k=1}^{m} \gamma_j(\mathbf{o}_i)_k \|o_{ij} - W_{:k} - Mr_i\|^2 \qquad (10)$$

where we use the notation $W_{:k}$ to denote the kth column vector of matrix W. The objective function in (10) is quadratic in both W and M, hence the optimization problem is jointly convex.

Here we propose to use the block coordinate descent algorithm to optimize the above objective function. Specifically, in each iteration, we sequentially find the global optima solution for each column vector $W_{:k}$ in turn, given all the other parameters fixed. Then we fix the updated weight matrix W and find the optimal solution for matrix M. To do so, we first derive the optimal solution for each column vector $W_{:k}, \forall k \in [m]$. To simplify the notation, we define $\tau := \sum_{i=1}^{n} t_i$ to be the total number of observations from all the sequences. Construct the following diagonal matrix Γ_k:

$$\Gamma_k := \mathrm{diag}\left(\{\{\gamma_j(\mathbf{o}_i)_k\}_{j=1}^{t_i}\}_{i=1}^{n} \right) \in \mathbb{R}_+^{\tau \times \tau}$$

and stack all the column vectors $o_{ij}, \forall j, i$ into a matrix $O \in \mathbb{R}^{d \times \tau}$. Furthermore, for each r_i, we also repeat it for t_i times and stack all such vectors into a matrix $R \in \mathbb{R}^{p \times \tau}$. With these new definitions, we can compactly represent the minimization problem w.r.t. $W_{:k}$ as the following matrix optimization problem:

$$\min_{W_{:k}} \qquad \|(O - MR - W_{:k}\mathbf{1}_\tau^T) \Gamma_k\|_F^2, \qquad (11)$$

from which we can derive the optimal solution with analytic form:

$$\hat{W}_{:k} := \frac{(O - MR) \Gamma_k^2 \mathbf{1}_\tau}{\mathbf{1}_\tau^T \Gamma_k^2 \mathbf{1}_\tau}. \qquad (12)$$

Note that here we define the huge and sparse matrix Γ_k only for the purpose of mathematical clarity. In practical implementation we will never explicitly construct it. Instead, a column vector $\Gamma_k^2 \mathbf{1}_\tau$ will be constructed and used, which only takes $O(\tau)$ time and space.

Analogously, to solve for M, we can compactly rewrite the optimization problem w.r.t. M as:

$$\min_{M} \qquad \sum_{k=1}^{m} \|(O - MR - W_{:k}\mathbf{1}_\tau^T) \Gamma_k\|_F^2, \qquad (13)$$

and solve it in closed form as:

$$\hat{M} := \left(\sum_{k=1}^{m} (O - W_{:k}\mathbf{1}_\tau^T) \Gamma_k^2 R^T \right) \left(R \sum_{k=1}^{m} \Gamma_k^2 R^T \right)^{-1}. \qquad (14)$$

The optimal solution $\hat{M}$ can be computed in time $O(\tau dp + p^3)$.

3.2.3 Overall Learning Algorithm. The overall learning algorithm is an iterative method that can be understood as an EM algorithm tailored to capturing patients' static features and absorbing states. In each iteration of the algorithm, for each observation sequence, we first compute the forward and backward probabilities, from which we could obtain the current belief state of each time step in the sequence. This corresponds to the E-step of the algorithm where we compute the expected complete-data log-likelihood function given current model parameters. In the M-step, we apply the procedures derived in Sec. 3.2.1 and Sec. 3.2.2 to obtain updated model parameters. Like the traditional EM algorithm, each iteration is guaranteed to improve the marginal log-likelihood function of the observation sequences, and the algorithm terminates if a stationary solution is found. We summarize the proposed algorithm in Algorithm 1.

Each iteration of the algorithm takes time $O(\tau(m^2+dp)+p^3)$. Furthermore, since the weighted least square problem (10) is smooth and convex, the block coordinate ascent subprocedure in Algorithm 1 is guaranteed to find the global optimal solution of both W and M in each EM iteration. Hence each EM iteration in Algorithm 1 is guaranteed to increase the log-likelihood function, and the algorithm is guaranteed to converge.

3.3 DyCRS inference: Converting Hidden States to Complication Risk Score

Given the learned HMM $\mathcal{H}$ from the learning process, we start to construct a dynamic complication risk score. The traditional HMM's decoding task can only tell us the most likely hidden states, however, the hidden states are discrete instead of a real number to quantify the complication risk. Besides it is meaningless and abstract to show the patient is in low risk stage, a real number that represents the probability the patients will get to complication event will be more meaningful and straightforward. To build a risk score, we develop a pipeline to convert the discrete hidden states to the continuous complication risk score. At a high level, this pipeline consists of two steps: (*i*) **decoding step:** given a new patient vital signs and static features up to time t and learned HMM parameters, we first decode the hidden states distribution for the patient. (*ii*) **converting step:** then we convert the hidden states distribution to complication risk score using the learned transition relationship between hidden states and observed clinical worsening state (C).

The complication risk score $CRS_t(\mathbf{o}, r)$ at time step t is defined as the probability that the patient will end in the clinical worsening state (C) given her observation sequence $\mathbf{o}$ up to time t and static

features vector r. The score at time t is given by:

$$CRS_t(\mathbf{o}, r) := \sum_{k=1}^{m} p(s_t = k \mid o_{1:t}, r) \times p(s_\infty = C \mid s_t = k), \quad (15)$$

where $p(s_t = k \mid o_{1:t}, r)$ is decoded current hidden states distribution. $p(s_\infty = C \mid s_t = k)$ is the absorption probability, i.e., the probability that the patient will finally end in the absorbing clinical worsening state C starting from current hidden state.

Decoding Step: Given the new patients' vital signs and static features and learned HMM parameters, we first decode the current hidden states distribution $p(s_t = k \mid o_{1:t}, r)$ which is exactly the forward probability $\alpha_j(\mathbf{o})$ at time t. This term can be efficiently calculated by the forward algorithm.

Converting Step: We further converting the hidden states distribution to complication risk score using the absorption probability (starting from the current hidden state, the probability the patient will be finally absorbed in clinical worsening state C). The absorption probability can be derived from the estimated transition matrix in the following way:

With 2 absorbing states and 3 transient states in our model, the learned transition matrix T has the following canonical form:

$$T = \begin{array}{c} \\ D \\ L \\ M \\ H \\ C \end{array} \begin{array}{ccccc} D & L & M & H & C \\ \begin{pmatrix} 1 & T_{DL} & T_{DM} & T_{DH} & 0 \\ 0 & T_{LL} & T_{LM} & T_{LH} & 0 \\ 0 & T_{ML} & T_{MM} & T_{MH} & 0 \\ 0 & T_{HL} & T_{HM} & T_{HH} & 0 \\ 0 & T_{CL} & T_{CM} & T_{CH} & 1 \end{pmatrix} \end{array}. \quad (16)$$

Define $p_D := (T_{DL}, T_{DM}, T_{DH})^T$ and $p_C := (T_{CL}, T_{CM}, T_{CH})^T$ to be the vectors that contain transition probabilities from transient states L, M, H to absorbing states D and C, respectively. Denote the middle 3×3 sub-matrix in T as Q. Then the absorption probability $p(s_\infty = C \mid s_t)$ can be computed as follows:

$$p(s_\infty = C \mid s_t) = \sum_{j=0}^{\infty} p(s_{t+j} = C \mid s_t) = p_C^T \sum_{j=0}^{\infty} Q^j = p_C^T (I - Q)^{-1}.$$

Combined with the formula for forward probabilities, the final $CRS_t(\mathbf{o}, r)$ can be compactly represented and efficiently computed as follows:

$$CRS_t(\mathbf{o}, r) = p_C^T (I - Q)^{-1} \alpha_t(\mathbf{o}). \quad (17)$$

For a sequence of length t, the score $CRS_t(\mathbf{o}, r)$ can be computed in $O(tm)$ time. Furthermore, using the forward algorithm, the update of $CRS_{t+1}(\mathbf{o})$ from $CRS_t(\mathbf{o})$ could be finished in $O(m)$ time.

The score is computed in real-time in the sense that when a new vital sign is available, a new risk score will be calculated by the DyCRS system. When the risk score is higher than a predefined threshold, the DyCRS system will issue an alert. Otherwise, the system will wait for new observations to re-score the complication risk at the next time step. The overall system pipeline of inference in our DyCRS system is shown in Fig. 7.

Algorithm 1 Expectation-Maximization for DyCRS Learning

1: **while** not converged **do**
2:   // E-step:
3:   **for** $i = 1$ to n **do**
4:     Compute $\alpha_j(\mathbf{o}_i), \forall j \in [t_i]$ using the forward algorithm
5:     Compute $\beta_j(\mathbf{o}_i), \forall j \in [t_i]$ using the backward algorithm
6:     Compute belief state $\gamma_j(\mathbf{o}_i) \propto \alpha_j(\mathbf{o}_i) \circ \beta_j(\mathbf{o}_i), \forall j \in [t_i]$
7:   **end for**
8:   // M-step:
9:   Update initial distribution using (8)
10:   Update transition matrix using (9)
11:   // Block coordinate ascent:
12:   **while** W or M has not converged **do**
13:     **for** $k = 1$ to m **do**
14:       Fix all the other columns and update $W_{\cdot k}$ using (12)
15:     **end for**
16:     Fix W and update M using (14)
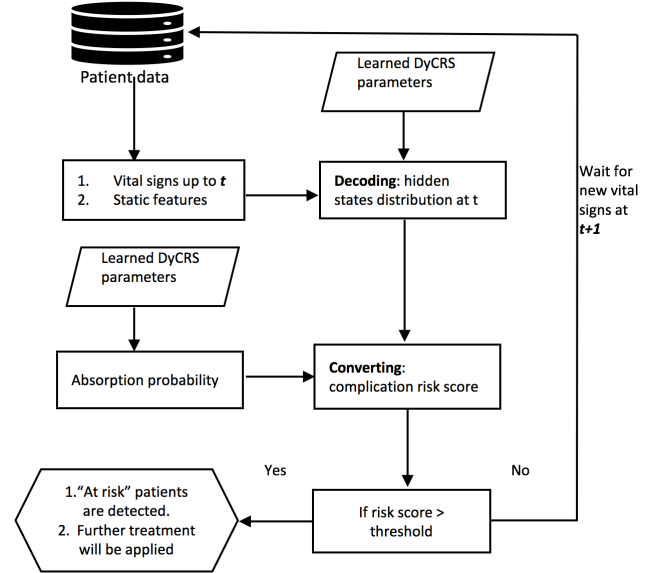17:   **end while**
18: **end while**

Figure 7: Flow chart of DyCRS real-time inference.

4 EXPERIMENTS

4.1 Experiment Setup

We compare the performance of the DyCRS framework with the state-of-the-art methods from both the medical and machine learning research communities. Baselines include four categories:

Machine learning real-time predictors. The sliding window approach is a common and widely used real-time prediction technique [20, 28, 40]. We compare with support vector machines with the radial based kernel (SVM), regularized logistic regression with ℓ_1 regularization (LASSO), random forest, AdaBoost, Bagging, and gradient boosting using the sliding window method. Specifically, we use a fixed-size window to filter the vital signs before the event occurs (clinical worsening or discharge). These fixed-length sequences

will be used as training data to learn the patient health conditions changes before the event occurrence. For the test set, we move the sliding window step by step and use each model to make a prediction, this predicted score will be used as a real-time complication score. The static features used in the DyCRS framework are also included to have a fair comparison. The size of the sliding window is a hyperparameter that needs to be tuned for each model. For the ensemble method like AdaBoost and Bagging, we use a decision tree as each component in the ensemble. We perform a grid search in {10, 50, 100} to find the optimal number of components for ensemble methods. To find the optimal size of the sliding window, we also conduct a grid search for each model in {12, 24, 36, 48} hours sequence prior to events.

Deep learning real-time predictor. We further compare with long short-term memory (LSTM), a deep learning technique for modeling temporal data [26, 39]. Specifically, the input at each time step is a 7-dimensional vector since we have 5 vitals signs associated with each time step and 2 static features. Inputs will be passed to the hidden layers. On top of the hidden layer, we use the sigmoid layer to squeeze the output to a risk score between 0 and 1. We perform a grid search to find the optimal number of hidden layers (H_l) and hidden size (H_s) ($H_l \in \{1, 2, 3\}$, $H_s \in \{20, 50, 80, 100\}$). Adam optimizer is used to optimize the parameters. We set the batch size as 32, the learning rate as 0.01. We train the LSTM for 200 iterations.

Pre-operative assessment: the state-of-the-art ACS NSQIP surgical risk calculator. ACS NSQIP is a widely used surgical risk calculator online web system in hospitals. It uses patients' static features to predict postoperative complications. ACS NSQIP has two stages: static features selection and estimation. Similar to the feature selection process described in the original paper of ACS NSQIP [9], we adopt recursive feature elimination to select features by recursively considering smaller and smaller sets of features. In the selection process, the estimator is trained on the original set of features and the importance of each feature is obtained via the coefficients of the estimator. The least important features are removed from the current set of features. The selection procedure is recursively repeated on the pruned set until the desired number of features to select is eventually achieved. For the estimator, it uses generalized linear mixed models that capture the random effects across hospitals and procedures. However, our dataset comes from one hospital and is only for elective Colectomy surgery, therefore the model can be reduced to logistic regression. As a result, we use logistic regression as the estimator directly. The number of features is a hyperparameter to be tuned, and we perform a grid search on {5, 10, 15, 20, 25, 30, 35} to find the optimal one.

Postoperative assessment: fever criterion used by clinicians in hospitals. Simple fever criterion is commonly used by clinicians as an indicator of potential complications once the surgery is over. Fever criterion is defined as: temperature $> 100.4^oF$. Therefore, when patients' postoperative temperature sequences after surgery and before event occurrence are greater than 100.4 ^oF, the patient will be labeled as a complication patient.

We use 5-fold stratified cross-validation. For each fold, we train the model using three parts, use one part for validation and the rest one part to report as the test score. The validation set is used for feature selection for the ACS NSQIP model and hyperparameter tuning for all the methods. We report the average test results.

4.2 Performance Metrics

To measure the performance of all real-time methods including the proposed DyCRS system and real-time baselines, we calculate risk score at each time step for all patients in our test set. Each patient will have a sequence of risk scores, and whenever the risk score exceeds the threshold, the targeted "at-risk" patients are found. The time when the risk score exceeds the threshold is denoted as T_{exceed}, and the event occurrence time (either clinical worsening or discharge) is denoted as T_{event}. We discuss the choice of threshold in more detail in the results part.

Accuracy metrics. We first look at the recall and precision rates.

$$\text{Recall} = \frac{\text{\# correctly predicted complication patients}}{\text{\# complication patients}} \quad (18)$$

$$\text{Precision} = \frac{\text{\# correctly predicted complication patients}}{\text{\# predicted complication patients}} \quad (19)$$

Earliness metric. Then we also want to see how early our real-time prediction can predict (T_{exceed}) ahead of clinician's decision (T_{event}). A larger earliness is desired since the sooner we can detect "at-risk" patients, the better it will help us to take measures to prevent it from causing greater adverse effects. We follow the [46] to define the earliness as:

$$\text{Earliness} = \mathbb{E}[T_{exceed} - T_{event}], \quad (20)$$

where the expectation is taken over all the correctly predicted complication patients.

Clinical significance metric. To further demonstrate the clinical significance of DyCRS system, we look at the number of false alarms per one correct alarm defined as:

$$\frac{\text{\# discharged patients predicted as complication patients}}{\text{\# correctly predicted complication patients}} \quad (21)$$

A lower number of false alarms per one correct alarm is desired, and it means we have more confidence to generate the alarms.

4.3 Results

Overall comparison. We use the area under the precision-recall curve (AUCPRC) to sweep all possible thresholds to have a fair and aggregated-level comparison. Note that we use a precision-recall curve instead of the receiver operating characteristic (ROC) curve since our data is slightly imbalanced. Previous studies show that the precision-recall curve is more informative than the ROC curve when evaluating on imbalanced datasets [36].

As shown in Table 1, our proposed DyCRS achieves the best AUCPRC (0.52) and significantly outperforms all the other machine learning, deep learning algorithms, and state-of-the-art ACS NSQIP calculator ($p < 0.01$) with 50.16% gain on average. We can also find that one-shot and static ACS NSQIP calculator works worse than DyCRS and other machine learning real-time predictors. This indicates that only using the patients' static features can not effectively detect postoperative complications. Instead, the postoperative vital signs can better reflect the actual physical condition of the patient after surgery, therefore are more informative to predict the postoperative complications than static features. To conclude, our proposed DyCRS, which can dynamically detect using both dynamic vital signs and static features, can be a better choice in terms of detection effectiveness.

Algorithms	AUCPRC	
	Mean	Std(%)
DyCRS	**0.52**	**1.55**
GradientBoosting	0.40	1.99
AdaBoost	0.42	2.83
Bagging	0.36	2.09
RandomForest	0.37	3.09
LASSO	0.35	2.15
SVM	0.34	2.69
ACS NSQIP	0.34	2.33
LSTM	0.25	1.43

Table 1: Overall Performance Comparison. Proposed DyCRS significantly outperforms other baselines ($p < 0.01$).

For real-time predictors, our DyCRS works better than all the machine learning real-time predictors based on the sliding window. This indicates the sliding window method using partial instead of full postoperative sequence are limited in measuring the long-term dependency and can only capture patients' health status changes within a fixed duration, hence some meaningful and representative clinical states that may be informative to the clinically worsening can be discarded, and as a result these methods may fail. For LSTM, the experimental results show that it always predicts trivial results like random guessing (e.g. always predict 0 rather than 1), which may due to the fact that it is difficult to train a deep learning model using sparse medical data. Therefore, we will exclude the LSTM in the following analysis.

Practical implications: trade-off between precision and recall rate. To show the practical implications, we further plot the trade-off between precision and recall rate in Fig. 8. On the one hand, for any specific recall rate, DyCRS achieves a significantly higher precision rate than benchmark methods. This means that if we fix the number of complication patients to identify, DyCRS will generate much less false alarms. For example, fixing a recall of 40%, DyCRS achieves a precision rate of 60%, which is 15.4% better than the best baseline (AdaBoost). This indicates DyCRS is more accurate with generated complication alarms and thus could reduce alarm fatigue [38]. On the other hand, for any specific precision rate, the DyCRS system can detect significantly more postoperative complications patients.

The common postoperative assessment used by clinicians in the hospital, e.g., the fever criteria, does not work well (with a recall of 43% and precision of 32%). For the same recall rate of 43%, our DyCRS achieves a 92% higher precision rate than it. In fact, the fever criteria are worse than both ACS NSQIP and real-time predictors. This indicates that fever criteria alone are not a good indicator of postoperative complications.

Earliness and clinical significance. How "early" can these algorithms predict is another important aspect to measure. As shown in Fig. 9, DyCRS can predict 15hrs55mins earlier than the clinician's diagnosis with a recall of 60%, and DyCRS is 3hrs56mins earlier than other real-time predictors on average. This is also true for recall of 50% and 70%. The earliness of DyCRS will provide clinicians with an opportunity to take treatment actions before patients
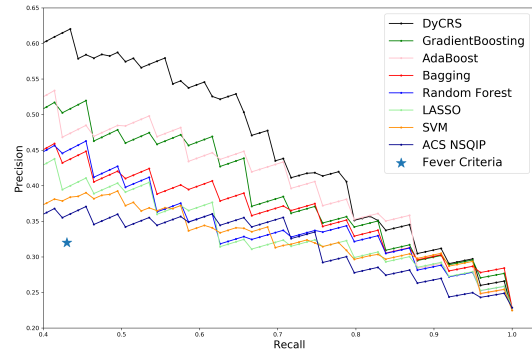

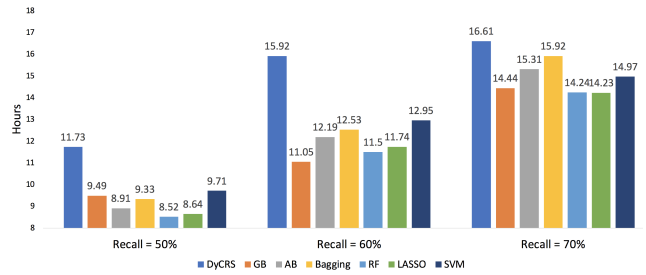**Figure 8: Trade off between precision and recall rate. DyCRS consistently outperforms other baselines.**

Figure 9: Earliness to measure how early can different algorithms predict prior to clinicians decision. GB, AB, RF represents gradient boosting, AdaBoost, random forest.
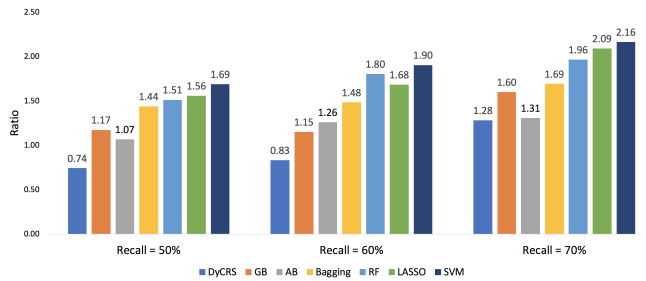

Figure 10: Clinical significance: the number of false alarms per one correct alarm. GB, AB, RF represents gradient boosting, AdaBoost, random forest.

deteriorate, and this will potentially reduce the adverse events and save healthcare costs.

To further show the clinical significance of DyCRS, we present the number of false alarms per one correct alarm in Fig. 10. For recall of 50%, DyCRS will generate only 0.74 false alarms per correct alarm. As a comparison, most other baselines will generate more than double false alarms as that of DyCRS. This indicates DyCRS is more confident about its alarms and potential will reduce the alarm fatigue [38] and improve hospital resource utilization as compared to other baselines.

Interpretability: consistent with previous medical postoperative complications studies. The estimated mean value of each

Table 2: Estimated Mean of Vital Signs Conditioned On Three Hidden States. WBC: white blood cell counts; BP: blood pressure. *L*, *M*, *H*: **low, medium, high risk states.**

Vital Signs	*L*	*M*	*H*	Normal Range
WBC	8.37	9.30	13.02	4.5-11.0
(10^9 cells per liter)				
Systolic BP (mm Hg)	120.69	150.78	126.87	<120 over 80
Diastolic BP (mm Hg)	64.08	80.49	68.59	<120 over 80
Pulse rate (min^{-1})	77.26	85.37	103.02	60-100
Temperature (^oF)	98.45	98.43	98.84	97-99

vital sign conditioned on the three hidden statesare summarized in Table 2. We expect *L*, *M*, *H* to represent the increasing level of risk w.r.t. complications (e.g. low risk, medium risk, high risk). Compared with *L*, *M* is a higher risk level w.r.t complications, since most criteria of *M* are higher than those of *L*, this is supported by previous medical studies that increased postoperative vital signs are associated with increased postoperative complications risk [25]. Specifically, systolic blood pressure and diastolic blood pressure exceed the normal range (less than 120 (systolic BP) over 80 (diastolic BP)). Compared to *M*, the average white blood cell counts conditioned on high risk state is around 13.02, which significantly exceeds the normal range of 4.5-11.0. This indicates potential infection, stress, inflammation, trauma, allergy, or certain diseases. This estimation is supported by previous postoperative complication medical studies that an abnormal postoperative white blood cell counts indicate the patients are at great risk of infectious complications [21, 24]. We also observe that the average pulse rate conditioned on *H* is 103.02, which is higher than the normal range of 60-100. Thus, our DyCRS can extract interpretable patients' stages, which are consistent with previous medical postoperative complication studies.

We also present the estimated transition matrix in (22). We find that the low risk states (*L*) have the highest probability of transfer to discharge event (*D*) compared to the other two intermediate-level risk states. This further indicates the low risk state has the smallest risk level among all the hidden states. On the other hand, high risk states have the highest probability to transfer to the clinical worsening event (*C*). This strengthens the increasing risk severity from low to high risk states in addition to the above estimated mean of vital signs conditioned on hidden states.

$$T = \begin{array}{c} D \\ L \\ M \\ H \\ C \end{array} \begin{pmatrix} 1 & 0.043 & 0.033 & 0.013 & 0 \\ 0 & 0.919 & 0.026 & 0.014 & 0 \\ 0 & 0.027 & 0.920 & 0.026 & 0 \\ 0 & 0.004 & 0.014 & 0.932 & 0 \\ 0 & 0.008 & 0.007 & 0.016 & 1 \end{pmatrix}. \quad (22)$$

5 RELATED WORK

Postoperative complication risk stratification. Most prior research has focused on using preoperative risk factors to stratify postoperative complication risks [9, 18]. Some work focus on developing procedure-specific surgical risk predictors(colorectal surgery [8], bariatric surgery [16], etc.). Others create universal postoperative complication risk calculators, which are procedure-free and can

capture the random effects of different hospitals [5]. These complication risk calculators are static and one-shot predict, Thus lack dynamic to capture the patients' changing postoperative health conditions. Different from them, our DyCRS is a dynamic complication risk scoring framework that can monitor the patient's complication risk in a real-time manner.

Real-time prediction. Real-time predictions are prevalent in many healthcare applications, including disease progression modeling [27, 31, 43], critical care prognosis [1, 2, 42, 46], and readmission [4]. Most clinical real-time risk prediction are based on sliding window approach with machine learning algorithms, e.g., SVM [44], logistic regression [19], etc. They train the model on the fixed-size window before the occurrence of events, then get the real-time score by moving the sliding window step by step on test sets. However, these methods can only capture patients' health status changes within a fixed duration, thus can not model the entire postoperative trajectory before events. In this way, some meaningful and representative clinical states which may be informative to the events can be discarded. Besides, recent work has explored deep learning techniques (recurrent neural networks [7], Long-short term memory (LSTM) [26]) to make real-time predictions in healthcare application. However, these approaches often require a large amount of training data, which can be expensive or difficult to obtain for surgery patients due to the privacy issue. Furthermore, it is often very hard to interpret the predictions obtained from these approaches, which is extremely vital and relevant for healthcare applications. Different from them, our DyCRS can learn the interpretable and meaningful representation that is consistent with the previous postoperative complication studies in medical research.

Time-series classification. Postoperative vital signs are essentially time-series data, therefore our work is also related to time-series classification. [12] extracted behavioral rhythms from wearable time-series data and used the extracted features to predict readmission. [45] proposed to use association rule mining to model the relationship between behavioral features to provide better depression prediction performance. However, these works focus on outputting a label for each entire sequence. Our work studies a different setting: in the postoperative setting, the patients' vital signs arrive step by step, therefore our DyCRS aim to give the postoperative complication risk score in a real-time manner, i.e., calculate the probability that the patient will have a complication at each time step when the new vital signs arrive.

6 CONCLUSION AND FUTURE WORK

In this paper, different from the state-of-the-art ACS NSQIP risk calculator which is preoperative assessment using patients' static features, we propose a new dynamic postoperative complication risk scoring framework that can score a patient in real-time using five postoperative vital signs and patients' static features. We demonstrate the superior predictive ability of our proposed DyCRS system compared to machine learning and deep learning methods, and the state-of-care in the medical area including ACS NSQIP calculator and a fever criterion approach commonly used by clinicians. In addition to effectiveness, we also demonstrate that our model can extract the interpretable patients' stages which are consistent with

previous medical postoperative complication studies, therefore the clinicians can trust our model as compared to the "black-box" predictors. In future work, we are interested in evaluating the DyCRS with clinicians in actual usage settings.

ACKNOWLEDGMENTS

REFERENCES

[1] Ahmed M. Alaa, Scott Hu, and Mihaela van der Schaar. 2017. Learning from Clinical Judgments: Semi-Markov-Modulated Marked Hawkes Processes for Risk Prognosis. In *Proceedings of ICML*. 60–69.

[2] Ahmed M Alaa and Mihaela Van Der Schaar. 2018. A hidden absorbing semi-Markov model for informatively censored temporal data: learning and inference. *The Journal of Machine Learning Research* 19, 1 (2018), 108–169.

[3] John B Ammori, Shawn J Pelletier, Raymond Lynch, Joshua Cohn, Yasser Ads, Darrell A Campbell, and Michael J Englesbe. 2008. Incremental costs of post–liver transplantation complications. *Journal of the American College of Surgeons* 206, 1 (2008), 89–95.

[4] Ofri Ben-Assuli and Rema Padman. 2019. Trajectories of Repeated Hospitalizations and Frequent Readmissions of Chronic Disease Patients: Risk Stratification, Profiling, and Predictions. *forthcoming in MIS Quarterly* (2019).

[5] Karl Y Bilimoria, Yaoming Liu, Jennifer L Paruch, Lynn Zhou, et al. 2013. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. *Journal of the American College of Surgeons* 217, 5 (2013), 833–842.

[6] Jeff A Bilmes et al. 1998. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. *International Computer Science Institute* 4, 510 (1998), 126.

[7] Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. 2018. Recurrent neural networks for multivariate time series with missing values. *Scientific reports* 8, 1 (2018), 6085.

[8] Mark E Cohen, Karl Y Bilimoria, Clifford Y Ko, and Bruce Lee Hall. 2009. Development of an American College of Surgeons National Surgery Quality Improvement Program: morbidity and mortality risk calculator for colorectal surgery. *Journal of the American College of Surgeons* 208, 6 (2009), 1009–1016.

[9] Mark E Cohen, Clifford Y Ko, Karl Y Bilimoria, Lynn Zhou, Kristopher Huffman, Xue Wang, Yaoming Liu, Kari Kraemer, Xiangju Meng, Ryan Merkow, et al. 2013. Optimizing ACS NSQIP modeling for evaluation of surgical quality and risk: patient risk adjustment, procedure mix adjustment, shrinkage adjustment, and surgical focus. *Journal of the American College of Surgeons* 217, 2 (2013), 336–346.

[10] Amy S Collins. 2008. Preventing health care–associated infections. In *Patient safety and quality: an evidence-based handbook for nurses*. Agency for Healthcare Research and Quality (US).

[11] John P Collins. 2019. Measures of Clinical Meaningfulness and Important Differences. *Physical therapy* (2019).

[12] Afsaneh Doryab, Anind K Dey, Grace Kao, and Carissa Low. 2019. Modeling Biobehavioral Rhythms with Passive Sensing in the Wild: A Case Study to Predict Readmission Risk after Pancreatic Surgery. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (2019), 1–21.

[13] Anahita Dua, Sapan S Desai, Bhavin Patel, Gary R Seabrook, Kellie R Brown, Brian Lewis, Peter J Rossi, Michael Malinowski, and Cheong J Lee. 2016. Preventable complications driving rising costs in management of patients with critical limb ischemia. *Annals of vascular surgery* 33 (2016), 144–148.

[14] Radwa Elshawi, Mouaz H Al-Mallah, and Sherif Sakr. 2019. On the interpretability of machine learning-based model for predicting hypertension. *BMC medical informatics and decision making* 19, 1 (2019), 146.

[15] RS Evans. 2016. Electronic health records: then, now, and in the future. *Yearbook of medical informatics* 25, S 01 (2016), S48–S61.

[16] Jonathan F Finks, Kerry L Kole, Panduranga R Yenumula, Wayne J English, Kevin R Krause, Arthur M Carlin, Jeffrey A Genaw, Mousumi Banerjee, John D Birkmeyer, Nancy J Birkmeyer, et al. 2011. Predicting risk for serious complications with bariatric surgery: results from the Michigan Bariatric Surgery Collaborative. *Annals of surgery* 254, 4 (2011), 633–640.

[17] Mohamed M Ghoneim and Michael W O'Hara. 2016. Depression and postoperative complications: an overview. *BMC surgery* 16, 1 (2016), 5.

[18] Prateek K Gupta, Himani Gupta, Abhishek Sundaram, Manu Kaushik, Xiang Fang, Weldon J Miller, Dennis J Esterbrooks, Claire B Hunter, et al. 2011. Development

[19] and validation of a risk calculator for prediction of cardiac risk after surgery. *Circulation* (2011), CIRCULATIONAHA–110.

[19] Joyce C Ho, Cheng H Lee, and Joydeeph Ghosh. 2012. Imputation-enhanced prediction of septic shock in ICU patients. In *Proceedings of the ACM SIGKDD workshop on health informatics (HI-KDD12)*. 18.

[20] Irfan Ahmad Khan, Adnan Akber, and Yinliang Xu. 2019. Sliding Window Regression based Short-Term Load Forecasting of a Multi-Area Power System. *arXiv preprint arXiv:1905.08111* (2019).

[21] Daehwan Kim, Ho-Seong Han, Yoo-Seok Yoon, Jai Young Cho, Youngrok Choi, Jae Yool Jang, and Hanlim Choi. 2015. Postoperative white blood cell counts change after pancreatoduodenectomy: Early sign for pancreatic fistula. *Korean Journal of Clinical Oncology* 11, 2 (2015), 95–100.

[22] Hilal Maradit Kremers, Sue L Visscher, James P Moriarty, Megan S Reinalda, Walter K Kremers, James M Naessens, and David G Lewallen. 2013. Determinants of direct medical costs in primary and revision total knee arthroplasty. *Clinical Orthopaedics and Related Research®* 471, 1 (2013), 206–214.

[23] Sunil Kripalani, Cecelia N Theobald, Beth Anctil, and Eduard E Vasilevskis. 2014. Reducing hospital readmission rates: current strategies and future directions. *Annual review of medicine* 65 (2014), 471–485.

[24] Gudrun Lamm, Johann Auer, Thomas Weber, Robert Berent, Cheung Ng, and Bernd Eber. 2006. Postoperative white blood cell count predicts atrial fibrillation after cardiac surgery. *Journal of cardiothoracic and vascular anesthesia* 20, 1 (2006), 51–56.

[25] Geoffrey K Lighthall, Sharmin Markar, and Robert Hsiung. 2009. Abnormal vital signs are associated with an increased risk for critical events in US veteran inpatients. *Resuscitation* 80, 11 (2009), 1264–1269.

[26] Zachary C Lipton, David C Kale, Charles Elkan, and Randall Wetzel. 2015. Learning to diagnose with LSTM recurrent neural networks. *arXiv preprint arXiv:1511.03677* (2015).

[27] Yu-Ying Liu, Shuang Li, Fuxin Li, Le Song, and James M Rehg. 2015. Efficient learning of continuous-time hidden markov models for disease progression. In *Advances in neural information processing systems*. 3600–3608.

[28] Ladan Mozaffari, Ahmad Mozaffari, and Nasser L Azad. 2015. Vehicle speed prediction via a sliding-window time series analysis and an evolutionary least learning machine: A case study on San Francisco urban roads. *Engineering science and technology, an international journal* 18, 2 (2015), 150–162.

[29] Montha Na Narong, Somchit Thongpiyapoom, Nonglak Thaikul, Silom Jamulitrat, and Nongyao Kasatpibal. 2003. Surgical site infections in patients undergoing major operations in a university hospital: using standardized infection ratio as a benchmarking tool. *American Journal of Infection Control* 31, 5 (2003), 274–279.

[30] Fouzia F Ozair, Nayer Jamshed, Amit Sharma, and Praveen Aggarwal. 2015. Ethical issues in electronic health records: A general overview. *Perspectives in clinical research* 6, 2 (2015), 73.

[31] Brenden K Petersen, Michael B Mayhew, Kalvin OE Ogbuefi, John D Greene, Vincent X Liu, and Priyadip Ray. 2018. Modeling sepsis progression using hidden Markov models. *arXiv preprint arXiv:1801.02736* (2018).

[32] Valentin Popov, Alesha Ellis-Robinson, and Gerald Humphris. 2019. Modelling reassurances of clinicians with hidden Markov models. *BMC medical research methodology* 19, 1 (2019), 11.

[33] Michael John Pritchard. 2012. Pre-operative assessment of elective surgical patients. *Nursing standard* 26, 30 (2012).

[34] Lawrence R Rabiner. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* 77, 2 (1989), 257–286.

[35] Katarzyna Rutkowska, Maciej Przybyła, and Hanna Misiołek. 2013. Health-care associated infection in the newly-opened intensive care unit. *Anaesthesiology intensive therapy* 45, 2 (2013), 62–66.

[36] Takaya Saito and Marc Rehmsmeier. 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one* 10, 3 (2015), e0118432.

[37] Andrew J Schoenfeld, Leah M Ochoa, Julia O Bader, and Philip J Belmont Jr. 2011. Risk factors for immediate postoperative complications and mortality following spine surgery: a study of 3475 patients from the National Surgical Quality Improvement Program. *JBJS* 93, 17 (2011), 1577–1582.

[38] Sue Sendelbach and Marjorie Funk. 2013. Alarm fatigue: a patient safety concern. *AACN advanced critical care* 24, 4 (2013), 378–386.

[39] Martin Sundermeyer, Ralf Schlüter, and Hermann Ney. 2012. LSTM neural networks for language modeling. In *Thirteenth annual conference of the international speech communication association*.

[40] Majid Vafaeipour, Omid Rahbari, Marc A Rosen, Farivar Fazelpour, and Pooyandeh Ansarirad. 2014. Application of sliding window technique for prediction of wind velocity time series. *International Journal of Energy and Environmental Engineering* 5, 2-3 (2014), 105.

[41] Wen Wang, Rema Padman, and Nirav Shah. 2018. Early Stratification of Patients at Risk for Postoperative Complications after Elective Colectomy. *arXiv preprint arXiv:1811.12227* (2018).

[42] Wen Wang, Rema Padman, and Nirav Shah. 2018. Early Stratification of Patients at Risk for Postoperative Complications after Elective Colectomy. *ArXiv* abs/1811.12227 (2018).

[43] Xiang Wang, David Sontag, and Fei Wang. 2014. Unsupervised learning of disease progression models. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 85–94.

[44] Jenna Wiens, Eric Horvitz, and John V Guttag. 2012. Patient risk stratification for hospital-associated c. diff as a time-series classification task. In *Advances in Neural Information Processing Systems*. 467–475.

[45] Xuhai Xu, Prerna Chikersal, Afsaneh Doryab, Daniella K Villalba, Janine M Dutcher, Michael J Tumminia, Tim Althoff, Sheldon Cohen, Kasey G Creswell, J David Creswell, et al. 2019. Leveraging Routine Behavior and Contextually-Filtered Features for Depression Detection among College Students. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–33.

[46] Jinsung Yoon, Ahmed Alaa, Scott Hu, and Mihaela Schaar. 2016. ForecastICU: a prognostic decision support system for timely prediction of intensive care unit admission. In *International Conference on Machine Learning*. 1680–1689.