Predicting the Onset of Tachycardia for Patients in Intensive Care Units

Lidan Mu
Machine Learning Department
Carnegie Mellon University
lmu@andrew.cmu.edu

Lujie Chen
Heinz College
Carnegie Mellon University
lujiec@andrew.cmu.edu

Joo Yoon
Department of Medicine
University of Pittsburgh
yoonjh@upmc.edu

Gilles Clermont
Department of Critical Care Medicine
University of Pittsburgh
clermontg@ccm.upmc.edu

Artur Dubrawski
Robotics Institute
Carnegie Mellon University
awd@cs.cmu.edu

Abstract

Tachycardia is a faster than normal heart rate at rest. In general, a resting heart rate over 100 beats per minute is accepted as tachycardia in adults [1]. While hospitalized patients can develop tachycardia from various benign clinical settings, some tachycardia events in the intensive care unit (ICU) could lead to cardiorespiratory instabilities such as shock and multi-organ hypoperfusion (lack of oxygenated blood flow) resulting in significant morbidity and mortality [2]. It is one of the most common types of instability in intensive care that happens before hypotension or organ failure. It is also one of the initial compensatory mechanisms to hypoxia and is one of the first measurable compensatory mechanism to shock. While most of ICUs are well-equipped to manage tachycardia, the event is not easily managed in majority of patients in a timely manner. This project aims to design and apply machine learning models that predict whether tachycardia will occur based on continuous bed-side monitoring data. It is intended to issue early warnings to enable earlier intervention and potentially prevent catastrophic events. The ultimate benefit will be increased survival rate among ICU patients.

1 Introduction

Tachycardia is a clinical phenomenon referring to a faster heart rate above the normal upper limit. In normal adult at rest, it is usually defined as heart rate greater than 100 beats per minute [1]. While hospitalized patients can develop tachycardia from various benign clinical settings, some tachycardia events in the intensive care unit (ICU) could lead to cardiorespiratory instabilities such as shock and multi-organ hypoperfusion (lack of oxygenated blood flow) resulting in significant morbidity and mortality [2]. Notably, tachycardia associated with shock and cardiorespiratory instability events has following characteristics, such as:

- one of the most commonly recognized findings in ICU patients;
- one of the initial compensatory mechanisms for intravascular volume depletion;
- one of the first measurable compensatory mechanisms to hypoxia or organ deficiency.

Critically ill patients with underlying cardiorespiratory comorbidities could have low reservoir in managing acute volume or pressure changes in their system with impaired elastance and contractility, and prone to develop tachycardia. In some cases, neurohormonal and metabolic triggers along with

vulnerable neural conduction system in myocardium can result in tachyarrhythmias (tachycardia with irregular ventricular response) that can be deleterious within a few minutes if untreated promptly [3].

Given the high risk of subsequent instability following tachycardia among critically ill patients, the development of prediction algorithm and associated risk score for tachycardia would enable timely identification and preemptive management of the high risk group, and enable better overall outcomes.

While most of ICUs are well-equipped to manage tachycardia, the event is not easily managed in majority of patients in a timely manner. This project aims to design and apply machine learning models that predict whether and how soon tachycardia will occur based on continuous bed-side monitoring data. It is intended to issue early warnings to enable earlier intervention and potentially prevent catastrophic events. There has been previous work on predicting atrial fibrillation [4][5][6], but no significant results have been achieved on the task of tachycardia detection and prediction.

The paper is organized in the following way: Section 2 gives details about the database used for training and testing the model, describes how controls and cases were selected and also what features were extracted for the prediction; Sections 3 and 4 talk about the experiments and analysis of the results; Section 5 gives an outline of possible future works to pursue along the path.

2 Methods

In this project, the main task is to identify records of time series that include signs of tachycardia and records that do not show these signs, and then perform classification based on these time series. Once we have trained a model that can classify time series with a reasonable accuracy, we would be able to apply the model on records of vital sign measurements and predict if tachycardia will occur in monitored patients. The data sets, labels and features used for the classification task, and the setup of the experiments are described below.

2.1 Datasets

The MIMIC II (Multiparameter Intelligent Monitoring in Intensive Care) Databases contain physiologic signals and vital signs time series captured from patient monitors, and comprehensive clinical data obtained from hospital medical information systems for Intensive Care Unit (ICU) patients. Data were collected between 2001 and 2008 from a variety of ICUs in a single tertiary teaching hospital [7]. The database contains the MIMIC II Clinical Database and the MIMIC II Waveform Database, and all databases are thoroughly de-identified (all personally identifiable information has been removed and all dates have been changed).

Records in the MIMIC II Clinical Database are clinical data recorded from bedside workstations as well as hospital archives [7], which are relatively discrete. The MIMIC II Waveform Database on the other hand, contains numeric time series of physiologic data, and hence can potentially carry timely information useful for the task of prediction. Records of continuous high-resolution physiologic waveforms and minute-by-minute numeric time series of physiologic measurements from the database are the major source of data used for this tachycardia prediction project.

2.2 Features

2.2.1 Feature Name Conversion

The temporal features extracted from the raw data are heart rate (HR), blood pressure (ABPSys, ABPDias, ABPMean), respiratory rate (RR) and arterial oxygen saturation (SPO2), which are all clinically significant features indicating the condition of a patient.

In the MIMIC II database, each record of time series has a slightly different naming convention. The records were processed according to the schema in Table 1, so that the naming of these features are consistent across all the time series used in our experiments.

Note that the use of symbols ABPs and NBPs are mixed together in some records, so the conversion was based on the fact that the values always follow the order ABPSys > ABPMean > ABPDias.

Feature Name		
raw data	processed data	
HR PULSE	HR	heart rate
ABPSys, NBPSys ABPDias, NBPDias ABPMean, NBPMean ABP.1, NBP.1 ABP.2, NBP.2 ABP.3, NBP.3 ABP, NBP	ABPSys ABPDias ABPMean	blood pressure
RESP RR	RR	respiratory rate
SPO2 SAO2	SPO2	arterial oxygen saturation

Table 1: Schema for feature name conversion in raw data.

2.2.2 Feature Extraction

Based on the 6 raw vital sign measurements extracted from the records of time series, further feature extraction step was performed. A set of the most common features used in time series classification were selected for the task of prediction over a 30-minute time window. The choice of the length of this time window will be explained later. The features are described in more details in Table 2.

2.3 Labels

There are a total of 2808 patients in the waveform data records, where each patient is associated with multiple records of time series. Records in the database have different frequencies, where some of them have recorded numerical values every 60 seconds, and some every 1 second. In order to keep the time granularity consistent, we only used records with time interval of 1 second. The way we defined controls and cases is described in the following sections.

2.3.1 Definitions

In each record of time series, a tachycardia event is defined as any time interval where the heart rate is at least 130 times per minute and a tachycardia episode is defined as a set of tachycardia events where every event is no more than 30 minutes apart, as shown in Figure 1.

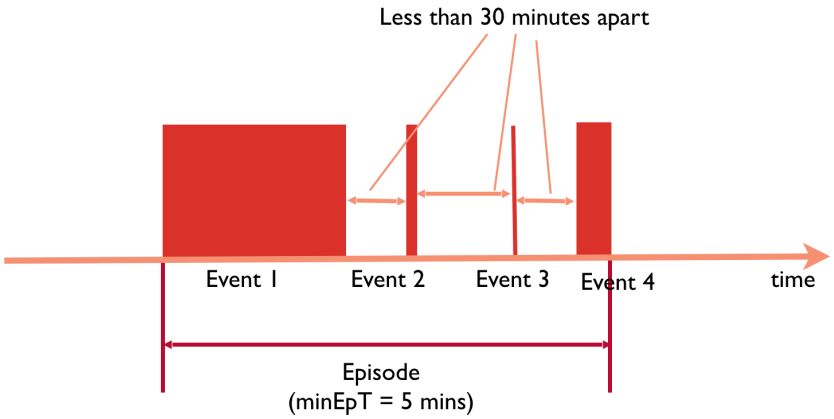

Figure 1: An example of a tachycardia episode.

Feature	Detail	Explanations
mean_spo2 mean_rr mean_hr mean_abpsys mean_abpdias mean_abpmean	mean values	
sd_spo2 sd_rr sd_hr	standard deviations	
reg_spo2 reg_rr reg_hr reg_abpsys reg_abpdias reg_abpmean	coefficient of first-order regression	
fft_spo2 fft_rr fft_hr	fast Fourier transform	
acs_spo2 acs_rr acs_hr	autocorrelation	The correlation of a signal with itself at different points in time.
aes_spo2 aes_rr aes_hr	approximate entropy	aes reflects the likelihood that similar patterns of observations will not be followed by additional similar observations.
ses_spo2 ses_rr ses_hr	sample entropy	
density_spo2 density_rr density_hr	density of the records	
last_5min_mean_spo2 last_5min_mean_rr last_5min_mean_hr	mean values in the last 5 minutes	The conditions of the patient in the last 5 minutes and last 10 minutes can deteriorate quickly, so the mean values and the coefficients of first-order regression reflect how fast the conditions worsen.
last_5min_reg_spo2 last_5min_reg_rr last_5min_reg_hr	coefficients of first-order regression in the last 5 minutes	
last_10min_mean_spo2 last_10min_mean_rr last_10min_mean_hr	mean values in the last 10 minutes	
last_10min_reg_spo2 last_10min_reg_rr last_10min_reg_hr	coefficients of first-order regression in the last 10 minutes	

Table 2: Features used for prediction.

We also define the duty cycle of a tachycardia episode E as

$$\text{duty cycle} = \frac{\sum_{\text{event}_i \in E} \text{duration of event}_i}{\text{total duration of } E}$$

which defines a concept of the density of a tachycardia episode. A tachycardia episode with higher duty cycle value therefore tends to consist of tachycardia events that are more persistent and/or closer to each other.

For cases where the episode does not last long enough or tachycardia events are merely occasional, these episodes are less significant as compared to cases where the episode persists for a long period of time. In our experiments, the former cases were omitted. We use a duty cycle of at least 0.1 and 5 minutes as the minimum duration of an episode to select tachycardia episodes for prediction.

2.3.2 Experiemnt I

During patient stays in the ICUs, there can be multiple tachycardia episodes based on our definition. One way to define the prediction problem is to select all tachycardia episodes as cases and any patient stay without any tachycardia episode as controls.

Since every single tachycardia episode is potentially dangerous for a patient, we can treat each episode as equally important when performing prediction. Using this definition of the experiment, we hope to differentiate patients who will develop tachycardia during their stay in the ICUs from those who will never have tachycardia. Following this definition, we selected 787 tachycardia episodes in total from all the records as the positive group.

Right after the admission of patients into the ICUs, the conditions of a patient might be unstable. Therefore, shorter stays would be less desirable to be included in the control group, in the sense that they would contain less information about the stabilized conditions of patients. At the same time, to keep the number of cases and controls roughly the same, setting a threshold on the minimum length of the record is also helpful. After comparing different choices of the threshold, a minimum of 4 hours' record is set as the threshold for selecting the controls, which gives us 707 records of control time series in total.

2.3.3 Experiment II

Another way to define the prediction problem is to select only the leading tachycardia episodes during each patient stay as cases and any patient stay without tachycardia episode as controls.

Although each tachycardia episode can lead to instabilities in ICUs, a patient will have a higher probability of developing subsequent tachycardia episode once the first episode occurs. So it will be clinically more useful to predict the first tachycardia episode than to predict any of them. Based on this definition, there were 240 tachycardia episodes found in all the time series as cases. Another 240 time series were randomly selected from the 707 records mentioned above as controls.

One observation is that the heart rate of a patient remains relatively higher than normal after the first tachycardia episode, so predicting the first tachycardia episode during a patient stay is in general a more difficult task as compared to predicting subsequent episodes. Therefore, models trained under this setting should perform at least the same when used to predict all tachycardia episodes, if not better.

2.3.4 Experiment III

We can also select both controls and cases from time series that contain some tachycardia episodes. But controls should be way ahead of any of those episodes and cases should be chosen just before the onset of tachycardia episode.

This is a slightly different task from previous definitions. Patients who will develop tachycardia might have abnormal heart rate during their entire stay in ICUs, so being able to predict how soon tachycardia will occur based on records of time series would also be useful when we are performing tachycardia prediction. In this way, it would be possible to give preemptive treatment to those patients who we suspect that will develop tachycardia before its onset.

In order to make sure that the control time series do not contain information from cases and also to have a reasonable amount of time series for our experiments, we required that the control part and case part must be at least 15 minutes apart. A total of 450 time series were selected, where half of them are cases and another half are controls.

2.4 Prediction Horizons

The records of time series span the whole stay of each patient in the ICU. In order to learn predictive models from this data, we prefer using relatively short interval of time to produce inputs, so that monitoring can begin shortly after admission. A 30-minute time window of the time series contains enough information of a patient based on empirical evidence.

When we are training the model, this 30-minute time window can be set right before the onset of tachycardia. But since tachycardia is just a faster heart rate than normal, heart rate in the 30-minute

time window right before the onset can be already higher. Models trained this way are very likely to pick up less information apart from features related to heart rate. Also, a model that can predict tachycardia long ahead of its onset would be more useful than a model that can only accurately make the prediction right before it happens.

Therefore, we trained several models with different horizons of prediction $n = 0, 1, ..., 30$ minutes – by setting this 30-minute time window $n = 0, 1, ..., 30$ minutes before the onset of tachycardia to select a reasonable gap between time of prediction and tachycardia. All models were validated using 10-fold cross validation, which specifically avoided assigning data from the same patient stay to both training and testing partitions in the same iteration.

2.5 Models

We used logistic regression with Lasso and random forest models to perform the task of classification.

2.5.1 Logistic Regression with Lasso

Logistic regression model is used here as baseline classifier to perform tachycardia prediction. With Lasso (or ℓ_1 regularization) [8], it can be robust to over-fitting and include feature selection.

2.5.2 Random Forest

Random forest [9] has been applied in several previous clinical researches. The model tends to be robust to over-fitting problems and do not expect linear features or even features that interact linearly as compared to regression models. Since missing records are common in the numeric data, the implementation of random forest model we used can handle missing values in the training data.

3 Results

3.1 Analysis

As shown in Figure 2 and Figure 3, the performance of random forest model systematically if not significantly outperforms the baseline classifier, logistic regression with Lasso. Note that the bars in the figures represent the confidence interval. Therefore, further analysis will focus on results given by random forest models.

Observing Figure 4 and Figure 5, with different choices of the parameter for horizon, we can observe that the accuracy and Area Under the Receiver Operating Characteristic Curve (AUC) decrease, as the horizon shifts away from the onset of tachycardia. Again the bars in the figures represent the confidence interval. This trend holds for all three different settings of experiments. Hence there is a tradeoff between high accuracy and making the prediction sooner.

The figures also confirm our intuition that predicting any tachycardia event is the easiest task, and the other two experiments that aim to predict the first tachycardia event always achieve lower accuracies and AUCs. By comparing the latter two experiments, it is obvious that the way we chose the control group greatly affected the performance of the model. Using cases and controls from the same time series is the most difficult one. But even with the last setting of the experiment, we still have a reasonable 10-fold cross validation accuracy of 0.8667 and AUC of 0.7290, enabling clinically useful forecasting of tachycardia even at early stages of ICU stays, which is very significant considering the current practice alternative of not having such indication at all.

3.2 Feature Comparison

By observing the top 15 features of the trained random forests models in Table 3, it is easy to see that as we moved the time window away from the tachycardia event, more features based on respiratory rate (RR) and arterial oxygen saturation (SPO2) were ranked higher in prediction. This confirms our intuition that the model would pick up less information apart from features related to heart rate if the time of prediction is set too close to the onset of tachycardia.

Figure 6 and Figure 7 show the accuracies and AUCs on experiment I with different choices of features. One set of models were trained using all 42 features, another set were trained using all

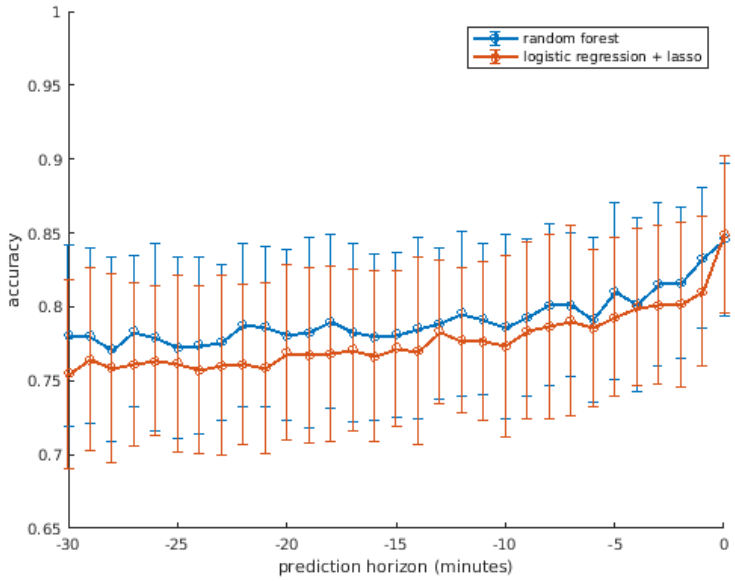

Figure 2: Accuracy of logistic regression and random forest trained for different horizons.

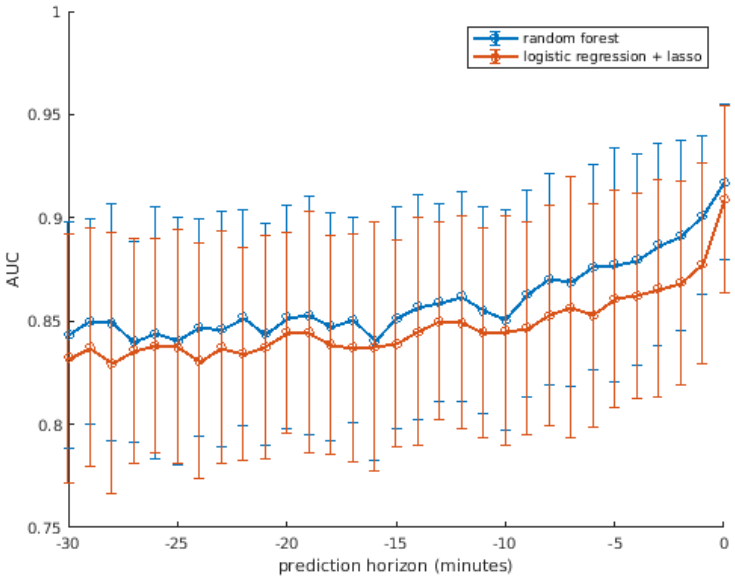

Figure 3: AUC score of logistic regression and random forest trained for different horizons.

features based on the heart rate raw features and the last set used all features except features based on heart rate. The heart rate features are clearly the most important set of features in the prediction task, and using this set of features ensures the performance of our model. Hence again, there is a tradeoff between high accuracy and practically useful generalization of the model.

3.3 Risk Trajectory

We extracted the same set of features over a moving 30-minute time window for each time series every 1 minute to form a time series of features. In each of the 10 cross validation folds, we used the

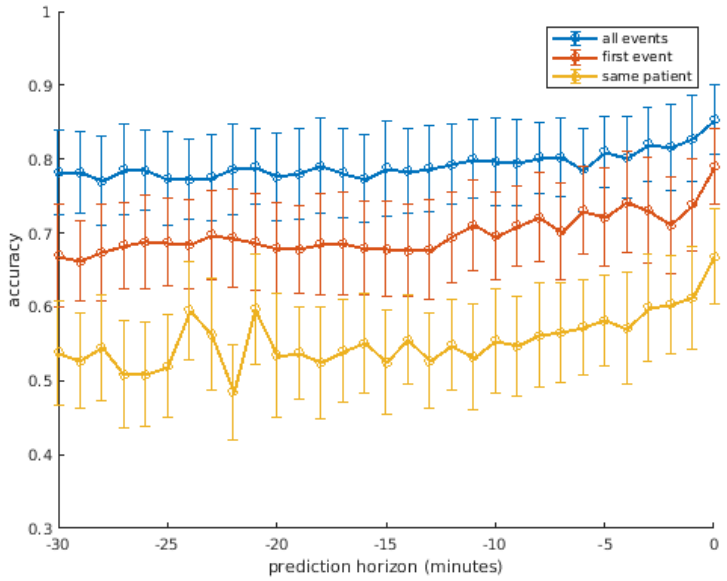

Figure 4: Accuracy of models trained for different experiments.

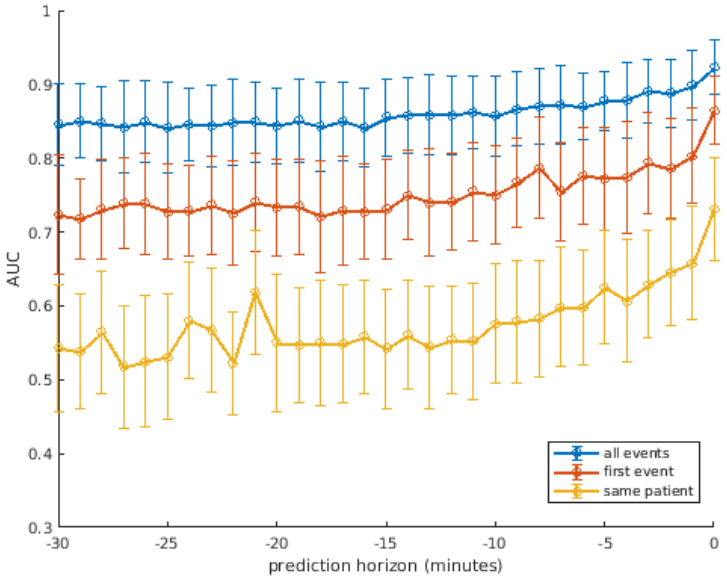

Figure 5: AUC score of models trained for different experiments.

trained model to compute tachycardia risk scores for time series in the testing partition. This can be viewed as estimating the likelihood of having tachycardia. We then separated the time series into two disjoint sets for controls and cases respectively, and derived the mean tachycardia risk scores at each time stamp for both sets to make comparisons. The risk trajectories are plotted in Figure 8, 9 and 10 (trained using all features), with the shaded areas representing the confidence interval. The x-axis denotes the time away from the onset of tachycardia in minutes and y-axis shows the mean risk score of having tachycardia.

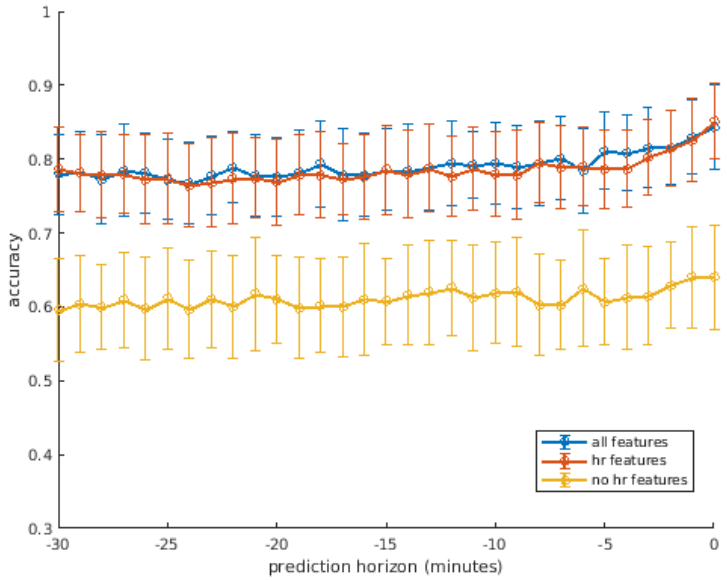

Figure 6: Accuracy of models trained for different choice of features.

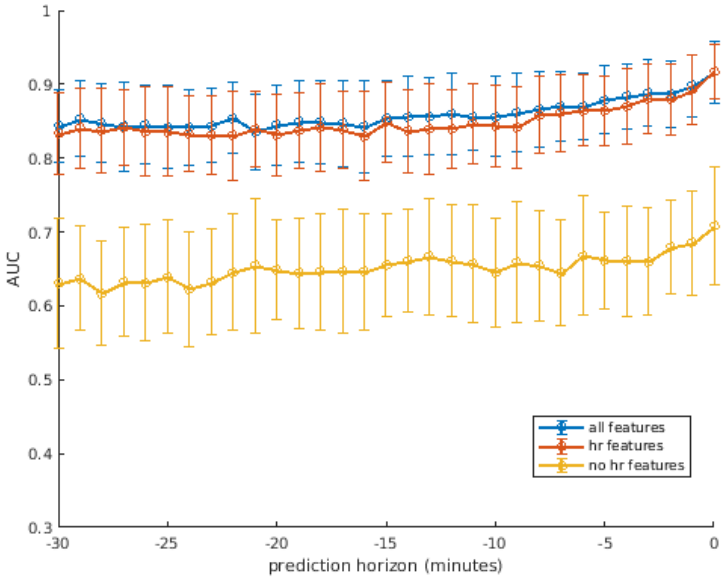

Figure 7: AUC score of models trained for different choice of features.

As we can see, for the first two settings of the experiment, the risk trajectories escalate as the time approaches the onset of tachycardia for cases, but they remain relatively unchanged for the control group. Also note that the curve for controls always lies below risk score of 0.5 and the curve for cases is above 0.5. The trajectories for models trained for various prediction horizons are all capable of showing this trend.

For the third setting of the experiment, since we selected controls to be long ahead of the cases, we can not align the time before the onset of the tachycardia for controls and cases the same way we did for the previous two settings. Hence we used the lift scores instead of the risk scores. Although score

0 min	10 min	20 min	30 min
aes_hr	fft_hr	aes_hr	aes_hr
fft_hr	**fft_rr**	fft_hr	fft_hr
last_10min_mean_hr	last_10min_mean_hr	**fft_rr**	last_10min_mean_hr
last_10min_reg_hr	**last_5min_mean_rr**	last_10min_mean_hr	**last_10min_mean_rr**
last_5min_mean_hr	mean_hr	last_10min_reg_hr	last_5min_mean_hr
last_5min_reg_hr	**mean_rr**	last_5min_mean_hr	**last_5min_mean_rr**
mean_hr	sd_hr	**last_5min_mean_rr**	mean_hr
mean_rr	ses_hr	mean_hr	sd_hr
sd_hr	**aes_spo2**	sd_hr	ses_hr
ses_hr	last_5min_mean_hr	ses_hr	**ses_spo2**
reg_hr	aes_hr	**mean_abpmean**	**mean_rr**
last_10min_mean_rr	**mean_abpmean**	reg_hr	**fft_rr**
fft_rr	**sd_spo2**	**mean_rr**	**reg_rr**
last_10min_reg_spo2	**last_10min_mean_rr**	last_5min_reg_hr	last_10min_reg_hr
last_5min_mean_rr	reg_hr	**last_10min_mean_rr**	last_5min_reg_hr

Table 3: Top 15 features with different horizons.

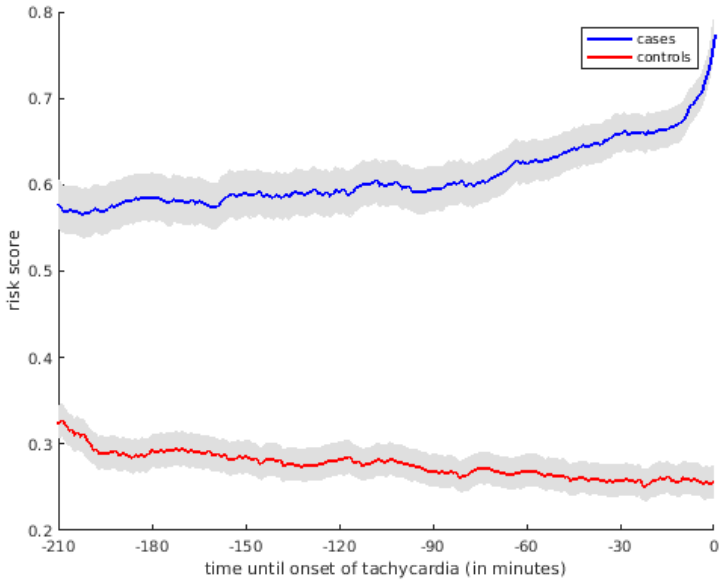

Figure 8: Risk trajectory of experiment I with prediction horizon ahead of the onset of tachycardia.

is very close to 1 at the beginning, the distinction becomes rather significant at the very end, which still leaves us about 15 minutes before the onset of tachycardia.

4 Conclusions

Based on the results, we can see that our models can identify who is going to have (the first episode of) tachycardia with high confidence. It could lead to potential high utility in triage to inform monitoring and health care resource allocation in the ICUs. Our models can confidently predict the incoming occurrence of tachycardia a substantial amount of time ahead of its onset in those patients who are at risk of developing it and could potentially trigger preemptive treatment.

Our results suggest a great potential impact in clinical settings, and can help improve quality of care and patient outcomes while mitigating costs of managing health of patients in intensive care.

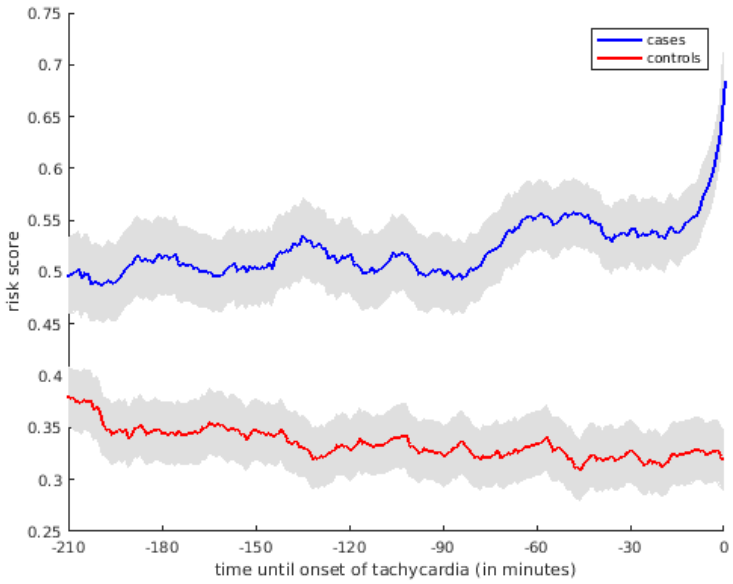

Figure 9: Risk trajectory of experiment II with prediction horizon ahead of the onset of tachycardia.

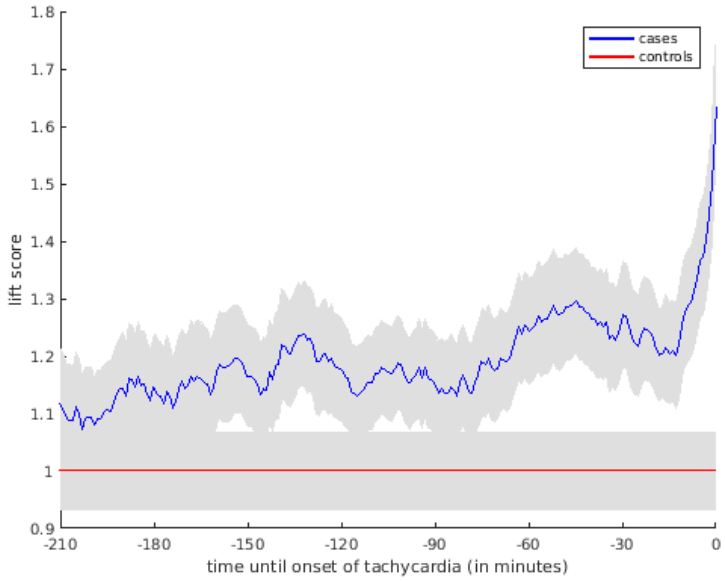

Figure 10: Risk trajectory of experiment III with prediction horizon ahead of the onset of tachycardia.

5 Future Work

In clinical research, some demographic features including age, gender or weight as well as medical history are also important drivers of patients' conditions. Given the fact the random forest model is able to handle these mostly categorical features, we plan to include the available demographic features into the model and compare the performance to the current models. They may potentially boost the accuracy and AUC scores.

Hypotension after tachycardia exposes patients to greater danger during their stay in the ICUs, so we will revise the selecting criteria for controls and cases groups to further predict co-occurrence of hypotension and tachycardia. This will be a valuable research topic related to tachycardia predication.

References

[1] Tachycardia. `https://en.wikipedia.org/wiki/Tachycardia`. Accessed: 2016-10-27.

[2] Sunghoon Park, Dong-Gyu Kim, Gee Young Suh, Woo Jung Park, Seung Hun Jang, Yong Il Hwang, Sang-Jin Han, Hyun Hee Jeong, Chang-Hoon Lee, and Ki-Suck Jung. Significance of new-onset prolonged sinus tachycardia in a medical intensive care unit: a prospective observational study. *Journal of critical care*, 26(5):534–e1, 2011.

[3] Djillali Annane, Véronique Sébille, Denis Duboc, Jean-Yves Le Heuzey, Nicolas Sadoul, Erik Bouvier, and Eric Bellissant. Incidence and prognosis of sustained arrhythmias in critically ill patients. *American journal of respiratory and critical care medicine*, 178(1):20–25, 2008.

[4] George Moody, A Goldberger, S McClennen, and S Swiryn. Predicting the onset of paroxysmal atrial fibrillation: The computers in cardiology challenge 2001. In *Computers in Cardiology 2001*, pages 113–116. IEEE, 2001.

[5] P De Chazal and C Heneghan. Automated assessment of atrial fibrillation. In *Computers in Cardiology 2001*, pages 117–120. IEEE, 2001.

[6] P Langley, D Di Bernardo, J Allen, E Bowers, FE Smith, S Vecchietti, and A Murray. Can paroxysmal atrial fibrillation be predicted? In *Computers in Cardiology 2001*, pages 121–124. IEEE, 2001.

[7] MIMIC II Databases. `https://physionet.org/mimic2/`. Accessed: 2016-10-24.

[8] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[9] Tin Kam Ho. Random decision forests. In *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*, volume 1, pages 278–282. IEEE, 1995.