Prediction of Deterioration in Critically Ill Patients with Heart Failure Based on Vital Signs Monitoring

Shengyu Zhang[1, 2], Kang Yang[2], Wenyu Ye[2], Haoyu Jiang[2], Xianliang He[2], Lei Wang[1], Yijing Li[2]

[1]Shenzhen Institutes of Advanced Technology, Chinese Academy of Science, Shenzhen, China
[2]Shenzhen Mindray Bio-Medical Electronics Co., Ltd., Shenzhen, China

Abstract

This study aims to develop a real-time machine learning model for acute heart failure onset based on vital signs in bedside monitoring. A group of 2284 patients were analyzed retrospectively from the Medical Information Mart for Intensive Care III database. We extracted various features building machine learning model. Extreme Gradient Boosting was used to develop the real-time prediction model. The validation on test set gave decent early warning performance. The model prediction can provide more timely notifications for doctors to perform better treatment for patients.

1. Introduction

Heart failure (HF), also known as congestive heart failure, is a type of sudden or chronic deterioration when the heart muscle does not pump blood as well as it should. When heart failure occurs, blood often backs up and fluid can build up in lungs, causing shortness of breath [1]. This is a severe situation for patients, having large possibility of inevitable consequences, like death. Clinical professions devote extensive research on treating heart failure, and found that an effective way is to handle deterioration early when heart failure just occurs [2].

Early alarms for heart failure, therefore becomes an issue of attention, but are very difficult to achieve for doctors in clinics, if only with the effort of health care workers. Usually, the early signs and symptoms of heart failure are subtle, including appearance of rapid or irregular heartbeat, shortness of breath with activity, fatigue or weakness, and so on. There is great possibility for health care workers to ignore these early signs and symptoms. If patient condition goes worse, severe signs or symptoms occurs, like chest pain, sudden shortness of breath and coughing up white or pink, foamy mucus, and so on. Clinical practice suggests that it is late to only notice severe changes of patient condition [3]. Also, in practical clinical scenarios, health care workers are often occupied by large number of tasks and unable to be at bedside all the time. They may miss the time window of finding early warnings for heart failure. Even more important, the recognition of early alarms for heart failure need comprehensive analysis and evaluation based on large amount of information, including monitoring data as well as the early signs and symptoms of patient. This issue is recognized as information overload in literatures [4], i.e., doctors are provided with too much information at their disposal. Unfortunately, for most doctors, an individual's efficiency in using information is not capable to analyse data of too large amount and deduce proper conclusions, which is especially hard for inexperienced doctors.

With the development of machine learning technology, it is possible to tackle the early-alarm difficulty, which hinders heart failure treatments in clinics. Machine learning technology can utilize the present overloading information for doctors to generate status indexes and early alarms of patients' deterioration.

Current study aimed to implemented an early warning machine learning system for heart failure. The system involved vital sign monitoring, which includes blood pressure, peripheral oxygen saturation (SpO2), heart rate (HR), respiratory rate (RR) and body temperature. These continuous monitoring input enabled machine learning to detect subtle changes that implies patients' heart-failure-related deterioration and therefore could provide timely warning before acute HF onset event. Multi-parameter waveform data in the Medical Information Mart for Intensive Care III (MIMIC III) database was used for system implementation [5].

2. Method

2.1. Dataset

We developed and tested our early warning method using MIMIC III database contained the multi-waveform data, including electrocardiogram (ECG), SpO2, invasive blood pressure (IBP) and respiratory waveform.

Figure 1 presents patient inclusion process. Subjects without multi-parameter waveform data or under age of 16 were excluded. ICD-9 codes about heart failure and

 ISSN: 2325-887X DOI: 10.22489/CinC.2022.264

ventilator interventions were used to identify acute heart failure (AHF), chronic heart failure (CHF) or non-heart failure (NHF) patients. CHF were not included for model development due to their indistinct status in terms of heart failure deterioration. AHF and NHF patients were further reviewed according to their electronic health record (EHR) information and waveform data, excluding cases with insufficient waveform data and resuscitation interventions (like high volume blood transfusion).

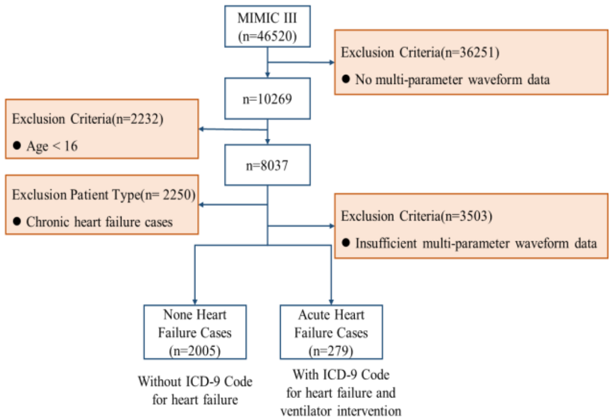

Figure 1. Patient inclusion and exclusion flow from the MIMIC III dataset

The acute heart failure onset event of AHF patients was preliminary obtained according to the ventilator operation recorded in the MIMIC III database, and then confirmed by experts. AHF patients all had one or more acute heart failure onset event, while NHF patients did not have any heart-failure-related acute event. We only used the first acute heart failure onset event of every AHF patient for model development.

2.2. System design

We utilized Mindray monitoring and analysis algorithm to pre-process the multi-parameter waveform data, resulting in vital signs with signal sampling rate of 1 Hz. We identified outliers in the time series data according to confidence interval of data points in a moving time window. Measurements exceed the range of confidence interval were excluded as outliers. Those outliers were replaced by their forward measurements, i.e., carry-forward imputation. The length of time window for processing outliers was set to 2 hours, with sliding step of 5 minutes. Missing values were also filled by carry-forward imputation.

After pre-processing of vital signs, including HR, systolic blood pressure (SBP), diastolic blood pressure (DBP), mean blood pressure (MAP), SpO2 and RR, we computed numbers of features to serve as input for machine learning algorithms. Raw features, differential features and sliding window-based statistical features were used in this study. The differential feature of time series data is the differences between current and its previous measurement. Sliding window-based statistical features involved a 12-hour time window sliding with step length of 1 hour and utilized various statistical method to obtain indexes of data in the time window. These statistical methods included mean, median, minimum, maximum, standard deviation and others (see Appendix for related feature description and further details) [6]. The above mention features were computed for each vital sign, i.e., each vital sign derived same number of features, except HR also involved heart-rate variability analysis [7].

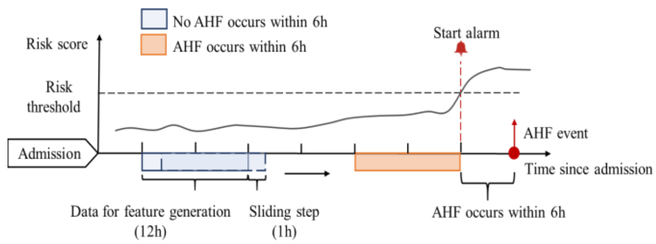

Figure 2. Illustration of real-time acute heart failure (AHF) onset event risk warning

With the features computed based on vital signs as input for machine learning, we developed prediction models based on extreme gradient boosting (XG-Boost). XG-Boost is a flexible machine learning model and can achieve good performance at representing nonlinear associations in data classifications. There are many implementations used XG-Boost for disease prediction in medical field. The output of XG-Boost, representing possibility of classification, can be used as an index for evaluation [3].

The acute heart failure event prediction was treated as classification of vital sign data segments. Patients' vital sign was divided into time series segments of 12 hours. Those data segments having acute heart failure onset within the next 6 hours was positive samples, while those data segments without onset in the next 6 hours was negative samples. Data segments of vital signs having acute heart failure onset event before the start of or during the period of segments were excluded, because of intervention impact on data. Each segment was processed to derived features for XG-Boost input, eventually predicting if there was AHF onset in the next 6 hours.

In sum, the proposed machine learning system was designed to predict the risk of acute heart failure onset within the next 6 hours, using vital sign data. Every hour, the system used vital sign data of last 12 hours to provide risk score. Comparing with predefined threshold, early alarms can be triggered. The design is shown as Figure 2. We also avoided over-fitting by three-fold cross validation in training. Ensemble approach was used to provide robust risk score, by averaging the outputs from three models generated in cross-validation process.

2.3. Experiment

A group of 2284 patients were analysed retrospectively from the MIMIC-III database. The start time of applying ventilator to patients was regarded as the acute HF onset, excluding scheduled operations of ventilator. Following our design, system training and testing were conducted. To evaluate the performance of the real-time prediction, AUROC, AUPRC, F-score and accuracy were computed, for both training and testing set.

To assess the impact of features on the prediction results, SHAP values were used to quantify the impact of each feature: positive or negative SHAP values indicate an increase or decrease of the prediction index.

3. Result

A total of 2284 patients were selected for system training and testing (279 AHF and 2005 NHF patients). After three-fold cross validation, a model combination, involving three sub models using XG-Boost, was derived. The average output of models was used as risk index. Prediction performance was evaluated by classification results of alarms for every hour, comparing to labels representing if there was acute AHF event in the next 6 hours. For training set, AUROC, AUPRC, F-score and accuracy were 0.973, 0.633, 0.478 and 0.901, respectively. While for test set, these performance indexes were 0.922, 0.401, 0.377 and 0.881. It was found from SHAP values that features of HR and RR provided high impact on the triggering of early warnings for acute heart failure events, like heterogeneity of HR or Hurst index of RR, shown in Figure 3. Figure 4 further presents Hurst index of RR had high positive impact when its value increase to one, indicating value close to one represented high risk for patients.

4. Discussion

Based on the MIMIC III database, a real-time and interpretable machine learning system, using vital signs as input, was developed. The system is suitable for bedside monitoring to predict acute onset of HF using vital signs. Accuracy is not the primary restriction for practical usage. If features used for prediction algorithms can not be obtained from vital signs monitoring, the early alarm for acute events will not be real-time, also restricted for clinical usage. The advantage of current solution is that, for every hour, our early warning system aimed to determine the risk of patient developing acute HF onset within the next 6 hours.

Besides real-time-processing, explanatory is also a critical issue for early warning. The past implementation usually can provide good prediction, but is unable to help doctors understand the relevant reason for the early warning. Recent research devotes extensive effort on explain the result of prediction. Current study was an attempt to make an explanatory early warning system. In our present envision for utilizing SHAP value for interpretation, if some features with high SHAP value present abnormal change when risk index give an early alarm, it is suggested that these features are associated with patient deterioration. For example, HR and RR can be viewed as features with high impact, which is in line with doctors' experience to pay attention to patients' HR and RR change [1].

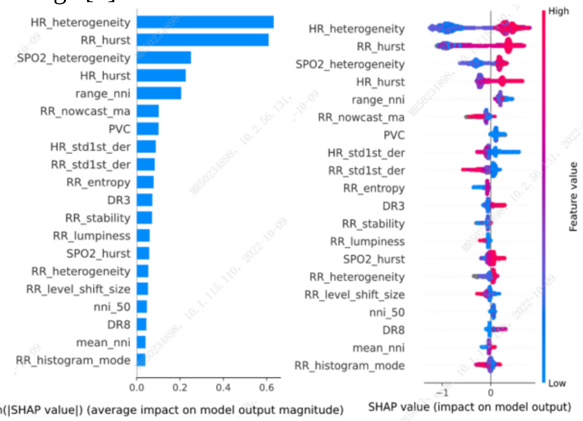

Figure 3. Summary about the impact of features. Twenty features with the highest mean absolute SHAP values are presented.

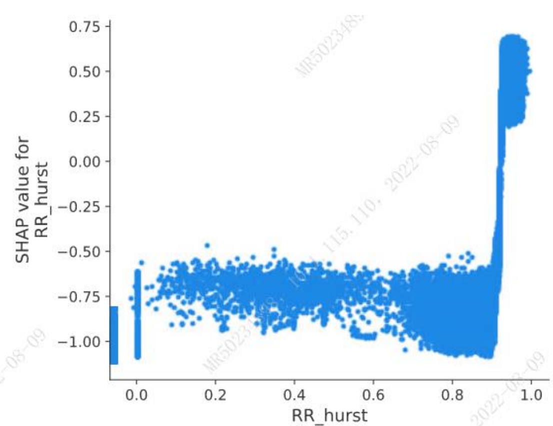

Figure 4. Scatter plot showing the association between feature value and SHAP value for Hurst index of respiratory rate (RR).

This study provides a concrete instance of advanced alarms for bedside monitoring. There is a trend to derive advanced alarms in research domain of bedside monitoring. The traditional alarms have drawbacks concerning the threshold. If threshold is difficult to reach, early warnings of patients' deterioration may be missed. But many false alarms will disturb clinical work if threshold too easy to reach. Existing products usually allow doctors to set their own thresholds, which is a compromise solution to this problem. The present machine-learning-based early warning system provided another possible approach for improving physician perception for monitoring alarms. By comprehensive analysis of time series data of vital signs, the alarm evidence will be more reliable than just alert based on a threshold.

There are several limitations to be considered in future research. First, alarm fatigue is a critical issue when early warning system put into practical use. In current research, we avoided the problem of alarm fatigue by decrease the alarm frequency. However, this design is inferior to continuous risk warning with high frequency, may providing alarms late for one hour if acute heart failure occurs just after last alarm time. Second, interpretation of feature requires a large population for feature-related research. Current study is still insufficient to generate precise advice to guide doctors for an in-depth vital-sign-based analysis of patients' condition. Large population research is still required. Third, our present phase of research is still retrospective, further study with actual clinical scenario is also required.

5. Conclusion

Current research proposed and validated a machine-learning algorithm for heart failure acute onset prediction. Our prediction index was built on vital signs from monitoring measurement and can provide early warning of heart failure event every hour, with decent accuracy. The SHAP value also gave insight on features for early warning, which can help interpreting algorithm prediction.

Appendix

Following table gives description of some features used in model.

Table 1. Description of feature used in model.

Feature Name	Description
entropy	Shannon entropy, the spectral density of time series
lumpiness	Variance of chunk-wise variances
stability	Variance of chunk-wise means
hurst	Long-term memory of time series
std1st_der	Standard deviation of first derivative of time series
heterogeneity	Heterogeneity of time series
spikiness	Variance of the leave-one-out variances of the remainder component
nowcast_ma	Mean of moving average time series data for short-term forecasting
level_shift_size	Size of the maximum mean value difference, between two consecutive sliding windows

Features mentioned in Table 1 are mainly based on Kats [6]. These features are computed for each vital sign, including HR, SpO2, RR, SBP, DBP and MAP. Detail description of features can be found in related website of 'Introduction to the tsfeatures package': https://cran.r-project.org/web/packages/tsfeatures/vignettes/tsfeatures.html.

Besides the above-mentioned features from Kats, derived features about heart-rate variability (HRV) analysis of ECG were also applied, like DR(X), representing R-peak intervals that increase in value with count more than X, and nni_50, representing number of consecutive R-peak intervals with differences greater than 50 milliseconds. For more information, refer to website about 'pyHRV': https://pyhrv.readthedocs.io/en/latest/.

References

[1] D. M. Henkel, M. M. Redfield, S. A. Weston, Y. Gerber, and V. L. Roger, "Death in Heart Failure," Circ. Hear. Fail., vol. 1, no. 2, pp. 91–97, Jul. 2008.
[2] S. Maji and S. Arora, "Decision Tree Algorithms for Prediction of Heart Disease," 2019, pp. 447–454.
[3] L. Yijing et al., "Prediction of Cardiac Arrest in Critically Ill Patients Based on Bedside Vital Signs Monitoring," Computer Methods Programs Biomed., vol. 214, p. 106568, Feb. 2022.
[4] I. Klerings, A. S. Weinhandl, and K. J. Thaler, "Information Overload in Healthcare: Too Much of a Good Thing?" Z. Evid. Fortbild. Qual. Gesundhwes., vol. 109, no. 4–5, pp. 285–290, 2015.
[5] A. E. W. Johnson et al., "MIMIC-III, A Freely Accessible Critical Care Database," Sci. Data, vol. 3, no. 1, p. 160035, Dec. 2016.
[6] Facebook Inc.: Kats for Python. [Online]. Available: https://facebookresearch.github.io/Kats/.
[7] A. Bauer et al., "Deceleration Capacity of Heart Rate as a Predictor of Mortality after Myocardial Infarction: Cohort Study," Lancet, vol. 367, no. 9523, pp. 1674–1681, May 2006.

Address for correspondence:

Yijing Li
Mindray Building
Keji 12th Road, High-Tech Industrial Park, Nanshan, Shenzhen, China
liyijing@mindray.com