Cross-Institutional Evaluation of SuperAlarm Algorithm for Predicting In-Hospital Code Blue Events

Randall Lee, MD, PhD[1], Ran Xiao, PhD[2], Duc Do, MD[3], Cheng Ding, MS[2], and Xiao Hu, PhD[2]
[1] *School of Medicine, University of California San Francisco,* [2] *School of Nursing, Duke University,* [3]*David Geffen School of Medicine, University of California Los Angeles*

Background

Bedside patient monitors are essential tools in acute care settings that provide timely information about patients' physiologic condition. However, current patient monitors are known to produce excessive alarms with majority of them being either false or non-actionable, leading to the well-known alarm fatigue issue. This creates an unsafe environment for patients as true alarms of advent events might be overlooked causing delayed or even missed opportunity for intervention. To tackle this issue, our previous studies proposed "SuperAlarm" algorithm that utilizes data mining and machine learning to find frequent combinations of monitor alarms that are predictive of in-hospital code blue (CB) events (Bai et al., 2015; 2017). The algorithm demonstrates high sensitivity in detecting code blue events while significantly reducing alarm frequency in an independent test set from the same healthcare institution. The present pilot study is our first attempt to evaluate the generalizability of SuperAlarm in different institutions, aiming to shed light on the implementation strategy of SuperAlarm algorithm.

Methods

Alarm data from bedside patient monitors located in critical care units were obtained from two healthcare institutions. Alarm data from institution A (InsA) include 412 code blue and 4020 matched control encounters (by time, diagnosis related group, age, gender and medical unit but without code blue calls) between 2013 and 2018. Alarm data from institution B (InsB) include 254 code blue and 2213 matched control encounters between 2010 and 2012. InsA data were divided into training and test set at 80%-20% split. The training set was used to mine SuperAlarm algorithm following the same framework as our previous studies (Bai et al., 2015; 2017). In brief, the Maximal Frequent Itemset (MAFIA) algorithm was adopted to find frequent combination of alarms (SuperAlarm patterns) preceding code blue events while seldomly occur in control encounters. The weighted average occurrence representation (WAOR) was used to integrate cumulative effect of SuperAlarm patterns over time, which then served as features to train a classifier (logistic regression with lasso regularization) to predict code blue events. All hyperparameters were determined through cross validation during the training process. Three metrics were calculated to evaluate SuperAlarm performance including sensitivity along lead time before code blue (Sen@L), alarm frequency reduction ratio (AFRR) and work-up to detection ratio (WDR). The derived algorithm was first tested on InsA test set to evaluate the internal institute performance, followed by external institute performance using all data from InsB as test set. Cross-institute performance evaluation was achieved by randomly selecting same number of encounters in InsB as in InsA test set 100 times via bootstrapping and statistically comparing the performance from both institutions.

Results

The final set of SuperAlarm patterns include in total 798 combinations with cardinality (number of alarms in a pattern) ranging from 2 to 7. Fig. 1A shows exemplar SuperAlarm patterns randomly selected from each cardinality. Fig. 1B presents differences in Sen@L between internal- (black circle) and external- (red circle) institutional test set, and bootstrap-based performance (red cross: mean, shaded areas: 95% CI). Performance on InsA and InsB test sets show significant difference in AFRR (0.931 ± 0.161 vs 0.950 ± 0.085, $p<0.05$) and WDR (7.064 ± 0.215 vs 8.614 ± 0.146, $p<0.05$).

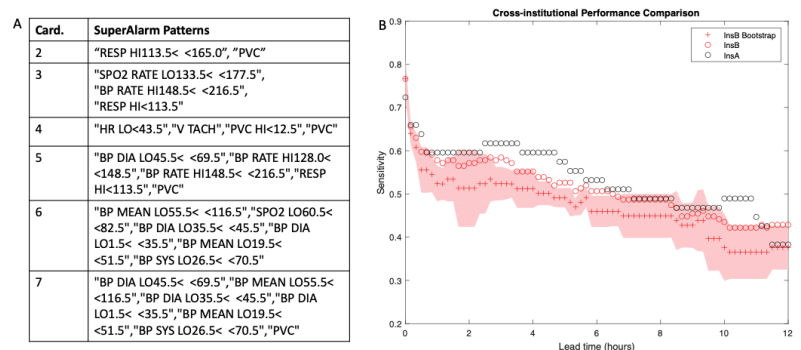| Card.	SuperAlarm Patterns
2	"RESP HI113.5< <165.0", "PVC"
3	"SPO2 RATE LO133.5< <177.5", "BP RATE HI148.5< <216.5", "RESP HI<113.5"
4	"HR LO<43.5", "V TACH", "PVC HI<12.5", "PVC"
5	"BP DIA LO45.5< <69.5", "BP RATE HI128.0< <148.5", "BP RATE HI148.5< <216.5", "RESP HI<113.5", "PVC"
6	"BP MEAN LO55.5< <116.5", "SPO2 LO60.5< <82.5", "BP DIA LO35.5< <45.5", "BP DIA LO1.5< <35.5", "BP MEAN LO19.5< <51.5", "BP SYS LO26.5< <70.5"
7	"BP DIA LO45.5< <69.5", "BP MEAN LO55.5< <116.5", "BP DIA LO35.5< <45.5", "BP DIA LO1.5< <35.5", "BP MEAN LO19.5< <51.5", "BP SYS LO26.5< <70.5", "PVC"

Figure 1 A. Exemplar SuperAlarm Patterns. B. Sen@L for internal and external institution performance.

Discussion

SuperAlarm patterns mined from InsA share similar arrangement as previous studies, such as coupling of arrythmia alarms with hemodynamic related alarms and combinations of parametric alarms that capture the trending information. Testing performance from both institutions shows large reduction in alarm frequency (over 90%) while maintaining above 60% sensitivity within one-hour lead time. It's worth noting lab test data that are found to improve sensitivity were not included due to the focus of the study (Bai et al., 2015). SuperAlarm presents significantly different performance from internal and external institution test sets as shown in Fig.1B. One plausible reason is the disparity in distribution of CB subtypes SuperAlarm between two institutions. Investigation of drill-down performance based on CB subtypes and clinical review of case-by-case acuteness will be our next steps to provide further insight to guide the implementation strategy of SuperAlarm.

Reference

Yong Bai, Duc Do, et al. Integrating monitor alarms with laboratory test results to enhance patient deterioration prediction. *Journal of Biomedical Informatics*, 53(11): 81–92, 2015.

Yong Bai, Duc Do, et al. Is the Sequence of SuperAlarm Triggers More Predictive Than Sequence of the Currently Utilized Patient Monitor Alarms? *IEEE Transactions on Biomedical Engineering*, 64(5): 1023–1032, 2017.