An Automated Method for Detecting Deterioration in Patient Vital Signs using Track-and-Trigger Observations

Sara Khalid, David A. Clifton, Lionel Tarassenko

Institute of Biomedical Engineering, University of Oxford, Oxford, UK, OX3 7DQ.
{Sara.Khalid, David.Clifton,Lionel.Tarassenko}@eng.ox.ac.uk

Abstract. Deterioration in patient condition is often preceded by physiological deterioration in vital-sign data. However timely detection of deterioration by using manual Track-and-Trigger monitoring is not always possible. We present an automated method for detection of deterioration based on Track-and-Trigger. We present results showing the proposed method outperforms the existing EWS scoring method which is based on clinical guidelines.

Keywords: Patient Monitoring, Early Warning Score, Track-and-Trigger

1 Introduction

It has been recognized that deterioration in patient condition is often preceded by physiological deterioration in vital-sign data [1], and that adverse events can be prevented by detecting this deterioration early; ultimately the burden on ICUs can be reduced.

However, timely detection of adverse events is not always possible because patient vital signs are typically monitored by clinical staff every four hours [2], with each nurse attending several patients. This often causes premonitory deviations in vital signs to go unnoticed. A recent study [3] on 326 in-hospital patients showed that only 7 out of 59 clinically adverse events were detected by clinical staff. Furthermore, studies [4] show that in nearly 25% of such cases where deteriorating trends in vital signs have been identified, they failed to elicit the necessary response from the clinical staff, possibly due to a lack of clearly defined response procedures.

1.1 Track-and-Trigger System

In the UK, the conventional system of assessing patient condition via vital signs is the track-and-trigger (T&T) system, which follows guidelines issued by the National Institute of Clinical Excellence (NICE) for assessing patient status [5]. Nurses record vital-sign observations at regular intervals on paper charts, enabling a patient's vital signs to be "tracked" through time. A T&T score is calculated based on an Early

Warning Score System (EWS), and if the score exceeds a certain threshold, the clinical team is "triggered" and the patient's condition is reviewed.

EWS systems have received mixed reviews in the literature, where some have criticised their effectiveness [6], while others have encouraged use of modified EWS in hospitals after modifications [7]. A common limitation of these systems is that they are based on manual checks performed by clinical staff, and these observations are carried out infrequently (e.g., once every four hours). Teams are thus activated based on these four-hourly scores. The infrequency of checks can significantly affects the predictive power of such systems.

1.2 Automated Alarm Generation

More recently, the focus has shifted towards developing "intelligent" early warning systems which are aimed at improving the clinical decision-making process by enabling automated monitoring of patient condition at regular intervals, and generating an alarm or "trigger" in the event of deterioration.

The problems associated with manual scoring systems are addressed by the use of automated monitoring systems. However, these systems cannot be used in wards where continuous monitoring is not carried out.

We therefore propose an automated early warning system for T&T monitoring that may be used in the absence of continuous monitoring. Further we aim to reduce the rate of false positives (i.e. the rate of false alerts) and false negatives (i.e. the rate of missed deterioration events) that are associated with the current T&T system in practice.

2 A Multivariate Bayesian Model for Patient Monitoring

We present a Bayesian model for the detection of deterioration in patient vital signs. A Bayesian approach is considered as it accounts for the uncertainty in low-frequency T&T data, while non-Bayesian approaches typically do not so. We consider multivariate data including heart rate and breathing rate for this work (and extension to include other vital signs in work-in-progress in this PhD.).

2.1 Data

Vital-sign data were collected from 200 patients recovering from upper-gastrointestinal surgery at the upper-GI ward of the Oxford University Hospitals NHS Trust, Oxford, UK[1]. Heart Rate (HR), Breathing Rate (BR), Systolic Blood pressure (SBP), Oxygen Saturation (SpO_2), and Temperature data were recorded by nurses on ward, typically every 4 hours on T&T paper charts. The data were then entered into an electronic database which was used for our analysis. 176 patients were considered as having normal recovery (average LOS 8.6 days) and placed in to a "normal-

[1] Approved by the Oxford Research Ethics Committee, OxREC No. 08/H0607/79.

recovery" group. 23 patients had post-operative emergency ICU admissions or died on ward, and were placed into an "adverse-event".

Bayesian Model Averaging. For multivariate data *D* belonging to the last 6 hours of a patient *i* (where $i = \{1,2,\ldots,176\}$) from the normal-recovery group, we define a multivariate model H_i as having a Normal-Wishart ($\mathcal{NW}$) distribution:

$$H_i \sim \mathcal{NW}(D|\boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i) \tag{1}$$

where the mean $\boldsymbol{\mu}_i$ and the precision $\boldsymbol{\Lambda}_i$ are the two parameters of the parameter-set $\boldsymbol{\theta}$. The evidence $P(D|H_i)$ of the i^{th} model H_i is the likelihood of the model which may be obtained after marginalising over the entire parameter space (or $\boldsymbol{\theta}$-space) over which the model exists:

$$P(D|H_i) = \int P(D|\boldsymbol{\theta}, H_i)\, P(\boldsymbol{\theta}|H_i)\, \mathrm{d}\boldsymbol{\theta} \tag{2}$$

Integrating over evidence for all possible models gives the marginal likelihood of data *D*, called the marginal likelihood, which may be expressed as:

$$P(D) = \int P(D|H_i)\, P(H_i)\, \mathrm{d}H_i \tag{3}$$

This concept is known as Bayesian Model Averaging (BMA) [8] and is based on the theory that averaging over all possible models allows better prediction regarding incoming data than a single model does. BMA is different from model combining in that while model combination tends to assign equal weight to all contributing models, BMA will automatically favour the model which has the highest evidence [9].

BMA Novelty Score. We defined a patient status index called the BMA score:

$$z = \log_e \frac{1}{P(D_{new})} = -\log_e P(D_{new}) \tag{4}$$

Here D_{new} is multivariate vital-sign data from a new patient. It is expected that "normal" data belonging to a recovering patient will take the lowest value of *z*, whereas "abnormal" data from patients with deteriorating vital signs are expected to score higher values of *z*.

We then introduced a threshold on *z*. We define this threshold, z', to be that which results in the smallest overall percentage of false positives and false negatives. An alert is generated if $z \geq z'$.

2.2 Comparison with existing EWS system

The proposed method was compared with the existing Early Warning System (EWS) used by nurses in the upper-GI ward to monitor a patient's vital signs manually. According to this system (shown in Table 8.1) which is based on current clinical guidelines, an individual vital sign scoring above 3, or an aggregate of individual scores above 4 generates an alert.

Table 1. The Early Warning Score (EWS) system in current practice at the upper-GI ward.

Vital Sign	*Score*	*3*	*2*	*1*	*0*	*1*	*2*	*3*
HR			≤ 40	41-50	51-100	101-110	111-129	≥ 130
BR		≤ 8			9-18	19-24	25-29	≥ 36
SpO_2		≤ 92			≥ 93			
SBP		≤ 90	91-99		101-179			≥ 180
TEMP			≤ 35		35.1-39.9			≥ 38

3 Results

Sensitivity was defined as the proportion of true positives in a window of data representing the last 24 hours from the adverse-event group. Specificity was defined as the proportion of true negatives in a window of data representing the last 24 hours from the normal-recovery group. The performance of the EWS method and the proposed method is shown in Fig. 1. It may be seen that the proposed method results in an "optimal" sensitivity and specificity of 40% at $z' = 0.5$, whereas the EWS method results in a very low false-positive rate (5.7%) which is encouraging, but also a very low sensitivity (8.1%), and is therefore far less accurate in detecting true deterioration events than the proposed method[2].

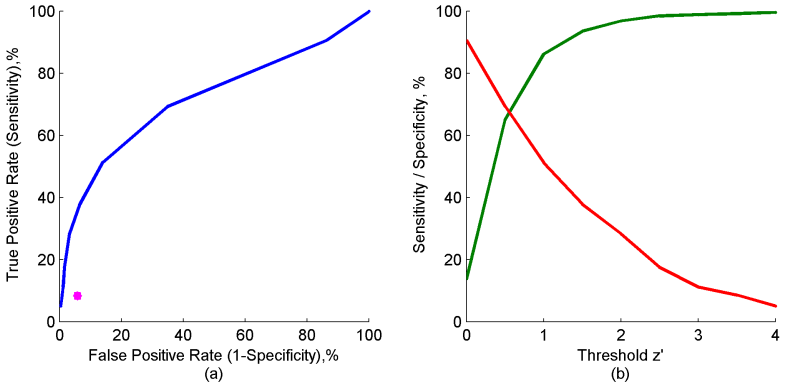

Fig. 1. ROC analysis. (a) For the proposed method, sensitivity is plotted against 1-specificity in blue for a range of thresholds *z*. Also shown in pink is the sensitivity and 1-specificity of the EWS system. (b) For the proposed method, the sensitivity and specificity are plotted against a range of *z*, in red and green, respectively. Optimal sensitivity and specificity is achieved at $z' = 0.5$.

[2] See Appendix A for explanation of poor performance of EWS method, which will be removed from the paper after review by LT and DC.

4 Conclusions

We presented a Bayesian model for detection of deterioration in multivariate vital-sign data. The proposed automated method was seen to perform with greater accuracy than the existing manual model in current practice.
The proposed method may therefore be developed for the automated monitoring of Track-and-Trigger data in hospitals.

References

1. M D Buist, E Jarmolowski, P R Burton, S A Bernard, B P Waxman, and J Anderson. Recognising clinical instability in hospital patients before cardiac arrest or unplanned admission to intensive care. A pilot study in a tertiary-care hospital. The Medical journal of Australia, 171(1):22–5, July 1999.
2. Julie T Sharpley and Janet C Holden. Introducing an early warning scoring system in a district general hospital. Nursing in critical care, 9(3):98–103.
3. Marilyn Hravnak, Leslie Edwards, Amy Clontz, Cynthia Valenta, Michael A Devita, and Michael R Pinsky. Defining the incidence of cardiorespiratory instability in patients in step-down units using an electronic integrated monitoring system. Archives of internal medicine, 168(12):1300–8, June 2008.
4. C Franklin and J Mathew. Developing strategies to prevent inhospital cardiac arrest: analyzing responses of physicians and nurses in the hours before the event. Critical care medicine, 22(2):244–7, February 1994.
5. Centre for Clinical Practice. Nice Guidelines: Acutely Ill patients in hospital. Technical report, National Institute of Clinical Excellence, 2007.
6. RWDuckitt, R Buxton-Thomas, JWalker, E Cheek, V Bewick, R Venn, and L G Forni. Worthing physiological scoring system: derivation and validation of a physiological early-warning system for medical admissions. An observational, population-based single-centre study. British journal of anaesthesia, 98(6):769–74, June 2007.
7. J Gardner-Thorpe, N Love, J Wrightson, S Walsh, and N Keeling. The value of Modified Early Warning Score (MEWS) in surgical in-patients: a prospective observational study. Annals of the Royal College of Surgeons of England, 88(6):571–5, October 2006.
8. Christopher M. Bishop. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer, 2007.
9. T Volinsky, E Raftery, A Hoeting, D Madigan. Bayesian Model Averaging: A Tutorial. Statistical Science, Vol. 14, No. 4, 382-417, 1999.

5 Appendix A

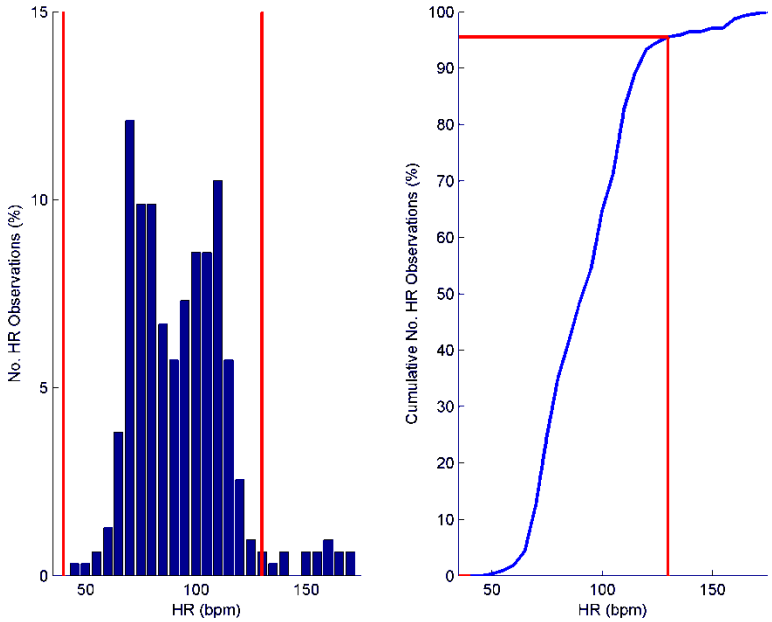

Fig. 2. Left: Distribution of HR observations for the adverse-event group recorded in the last 24 hours before event. The upper (130 bpm) and lower (40 bpm) alert thresholds for EWS method are shown in red. Right: The cumulative distribution of HR observations for the adverse-event group recorded in the last 24 hours before event. The EWS alert thresholds for HR are plotted in red.

From Fig. 2 it may be seen that the HR data from the adverse-event group used in performing the ROC analysis contains 90% data (as shown in Fig.2 b) which does not alert according to the EWS method. Fig. 3 shows the same distributions for the BR, and there too we see that ~ 90% BR data does not alert. We therefore expect the EWS method to be only 10% sensitive which is consistent with the results shown in Fig. 1a.

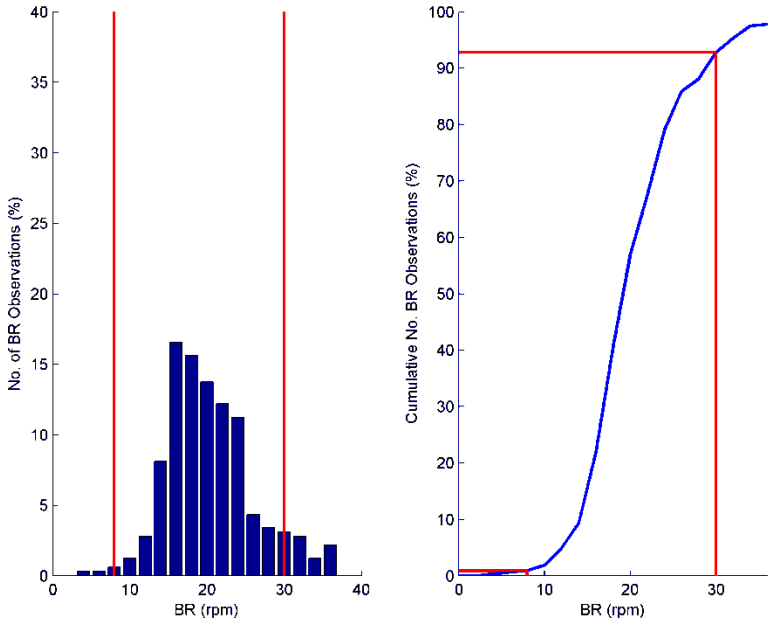

Fig. 3. Left: Distribution of HR observations for the adverse-event group recorded in the last 24 hours before event. The upper (130 bpm) and lower (40 bpm) alert thresholds for EWS method are shown in red. Right: The cumulative distribution of HR observations for the adverse-event group recorded in the last 24 hours before event. The EWS alert thresholds for HR are plotted in red.