International Journal of Medical Informatics 112 (2018) 15–20

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

An ensemble boosting model for predicting transfer to the pediatric intensive care unit

Jonathan Rubin[a,*], Cristhian Potes[a], Minnan Xu-Wilson[a], Junzi Dong[a], Asif Rahman[a], Hiep Nguyen[b], David Moromisato[b]

[a] Philips Research North America, Cambridge, MA, United States
[b] Cardon Children's Medical Center, Banner Health System, Mesa, AZ, United States

A R T I C L E I N F O

Keywords:
Early deterioration indicator
Machine learning
Pediatrics
Early warning systems

A B S T R A C T

Background: Early deterioration indicators have the potential to alert hospital care staff in advance of adverse events, such as patients requiring an increased level of care, or the need for rapid response teams to be called. Our work focuses on the problem of predicting the transfer of pediatric patients from the general ward of a hospital to the pediatric intensive care unit.

Objectives: The development of a data-driven pediatric early deterioration indicator for use by clinicians with the purpose of predicting encounters where transfer from the general ward to the PICU is likely.

Methods: Using data collected over 5.5 years from the electronic health records of two medical facilities, we develop machine learning classifiers based on *adaptive boosting* and *gradient tree boosting*. We further combine these learned classifiers into an ensemble model and compare its performance to a modified pediatric early warning score (PEWS) baseline that relies on expert defined guidelines. To gauge model generalizability, we perform an inter-facility evaluation where we train our algorithm on data from one facility and perform evaluation on a hidden test dataset from a separate facility.

Results: We show that improvements are witnessed over the modified PEWS baseline in accuracy (*0.77* vs. *0.69*), sensitivity (*0.80* vs. *0.68*), specificity (*0.74* vs. *0.70*) and AUROC (*0.85* vs. *0.73*).

Conclusions: Data-driven, machine learning algorithms can improve PICU transfer prediction accuracy compared to expertly defined systems, such as a modified PEWS, but care must be taken in the training of such approaches to avoid inadvertently introducing bias into the outcomes of these systems.

1. Introduction

Approximately 1–3% of pediatric patients admitted to the general ward of a hospital will be transferred to the pediatric intensive care unit (PICU) due to a deterioration in health [1]. Many guideline-based early warning score (EWS) systems that monitor a patient's state of health have been proposed to address this problem [2–4], as have data-driven approaches that rely on machine learned classifiers [5,6]. However, the majority of these systems have focused on adult populations, with less focus on pediatric patients where it is known that vital sign measurements, such as heart rate and respiration rate differ markedly in young children compared with adolescents and adults. Moreover, existing EWS systems aimed at young populations, such as Pediatric Early Warning Score (PEWS) [7,8], rely on manual spot check observations made by nursing staff, such as the capillary nail refill test, which means input into the system is subjective. An automated method that detects early deterioration in pediatric patients using physiologic vital sign information offers several advantages:

1. It ensures that patients that are in danger of deteriorating receive timely care and attention, thereby minimizing or avoiding harm to the patient due to the occurrence of a significant adverse event.
2. It does not rely on manually recorded information that is prone to subjective bias, such as capillary refill in the PEWS system.
3. It can inform the allocation of hospital resources.
4. It can reinforce the intuition of hospital care staff and act as further evidence when decisions about level of care are required to be made.

In this work, we present the development of an automated early deterioration algorithm for pediatric populations within a hospital's general ward. Our models accept a patient's age as input, as well as physiologic vital sign measurements. This information is used to make a

* Corresponding author.
E-mail address: Jonathan.Rubin@philips.com (J. Rubin).

https://doi.org/10.1016/j.ijmedinf.2018.01.001
Received 29 September 2017; Received in revised form 14 November 2017; Accepted 2 January 2018
1386-5056/

prediction about the likelihood of the patient transferring from the general ward to the PICU.

We compare two approaches based on ensemble boosting for creating transfer prediction models. The first relies on an *adaptive boosting* algorithm [9,10] that employs single level decision trees as its base classifier. The adaptive boosting procedure is altered to consider a patient's age for learning risk thresholds. The second approach constructs an ensemble of CART models (classification and regression trees) using *extreme gradient boosting* [11]. Finally, we combine the predictions of both the adaptive boosting and gradient boosting models into an ensemble and evaluate its performance. We compare the results of our *boosting*-based classifiers to a version of the Bedside PEWS Scoring System [8], which was modified based on available input data.

2. Cohort

Data was collected over a 5.5 year period from the electronic health records of two medical centers: Banner Thunderbird Medical Center and Banner Desert Medical Center. Encounters that occurred in the pediatric general ward(s) and pediatric intensive care unit were included in the datasets. Encounters where transfer occurred from the general ward to the PICU were determined using location and timestamp information from the electronic health record. All patients between the ages of 1 month and less than 20 years were included in the dataset. The study was approved by the Institutional Review Board (IRB) of Banner Health (Mesa, AZ, USA).

Table 1 summarizes the details regarding the number of unique patients and encounters, as well as patient demographic information. The values in Table 1 confirm that, for both facilities, there is a large imbalance between the number of pediatric encounters that resulted in transfer to the PICU compared to those that did not.

3. Feature selection, data preprocessing and splitting

3.1. Feature selection

We wished to construct a system that, given a snapshot of objective inputs, could make a determination about the likelihood of a patient being transferred to the PICU. The following features were selected to be used as inputs into the prediction model: 1. Heart Rate (HR); 2. O_2 Saturation (O_2); 3. Respiratory Rate (RR); 4. Temperature (Temp); 5. Diastolic Blood Pressure (dBP); 6. Systolic Blood Pressure (sBP); 7. Patient Age; 8. Pulse Pressure ($sBP - dBP$); 9. Approximate Mean Arterial Pressure ($2/3dBP + 1/3sBP$); and 10. Shock Index (HR/sBP).

The features listed above include direct vital sign measurements, age of the patient and three measurements derived from vital sign inputs. Laboratory values were also considered as input, as they have been included as features in adult deterioration indicators [12,5]. However, the extra stress induced in pediatric populations by performing blood draws meant that these inputs were generally collected less often and would likely be less available in practice, hence they were excluded as features. Spot check measurements such as Capillary Refill and Skin Color, originally included within PEWS systems, were excluded from the analysis due to their subjective nature.

Table 1
Patient encounters and demographic information per hospital facility.

	Desert		Thunderbird	
	Transferred	Non-transferred	Transferred	Non-transferred
Patients	305 (3.0%)	9982 (97.0%)	98 (1.9%)	5042 (98.1%)
Encounters	330 (2.6%)	12536 (97.4%)	102 (1.7%)	6005 (98.3%)
Average age	5.4 ± 5.7	6.1 ± 5.8	5.5 ± 6.1	6.3 ± 6.0
Gender				
– Female	130 (42.6%)	4491 (45.0%)	35 (35.7%)	2292 (45.5%)
– Male	174 (57.1%)	5273 (52.8%)	59 (60.2%)	2622 (52.0%)
– Missing	1 (0.3%)	218 (2.2%)	4 (4.1%)	128 (2.5%)

3.2. Data preprocessing

For each encounter that resulted in transfer to the PICU, feature values were retrieved from the electronic health record. Feature vectors were populated from clinical event measurements that occurred at least 2 h preceding the time of transfer and at most 8 h preceding transfer. The value used for each feature was the final clinical measurement recorded within the observation window, hence, each instance captured a snapshot of deterioration. Each instance that resulted in transfer was matched by a corresponding encounter that did not result in transfer. For non-transfer instances, a random 6 h observation window was selected and a snapshot of feature inputs consisting of the last recorded value in the observation window was used. In the case where no measurement was recorded for an input value within the 6 h observation window that feature's value was recorded as missing.

3.3. Data splitting

3.3.1. Training/cross-validation

Eighty percent of data from the Desert facility was used as training data. 10-fold cross validation was used to split this training data into separate folds. Choice of which hyperparameters to use for our models was based on maximizing the average cross-validation score over all 10 folds. Area under the receiver operating characteristic (AUROC) was used as the metric for optimization.

3.3.2. Testing

Twenty percent of data from the Desert facility was set aside as held-out test data. Stratified sampling was used to ensure an even class distribution.

One hundred percent of data from the Thunderbird facility was set aside as a separate held-out test-set, i.e. no encounter from the Thunderbird facility was used in model training/cross-validation. This decision was made to ensure that the final results obtained on the test-set accurately reflected generalizability between individual facilities. For the Desert dataset, it was further ensured that no patient who had any encounters in the training/cross-validation sets was included in the test set.

4. PICU transfer prediction algorithm

We compared two variants of boosting algorithms [13] for distinguishing between encounters that resulted in transfer to the PICU versus those that did not. Both algorithms were required to gracefully deal with missing feature values, as our dataset consisted of instances where certain vital sign information was missing and future deployment of such a system would require effective handling of missing information.

We wish to learn a model, $F_m(x) = y$, by recursively constructing baseline ("weak") classifiers, $h(x)$, fit to a specified loss function, $L(y, F(x))$. Beginning with an initial model $F_0(x)$, the final model, $F_m(x)$, is defined recursively by combing the predictions of the previous model, $F_{m-1}(x)$, with $h(x)$.

$$F_m(x) = F_{m-1}(x) + \alpha h(x), \quad m \geq 1 \tag{1}$$

where α is a scaling factor and m is the total number of baseline classifiers to fit.

4.1. Adaptive boosting

We first train an adaptive boosting classifier that seeks to add baseline classifiers, $h(x)$, that will minimize an exponential loss