Toward a Two-Tier Clinical Warning System for Hospitalized Patients

Authors: Gregory Hackmann, Minmin Chen, Octav Chipara, Chenyang Lu, Yixin Chen, Marin Kollef, and Thomas C. Bailey

Clinical study has found early detection and intervention to be essential for preventing clinical deterioration in patients at general hospital units. In this paper, we envision a two-tiered early warning system designed to identify the signs of clinical deterioration and provide early warning of serious clinical events. The first tier of the system automatically identifies patients at risk of clinical deterioration from existing electronic medical record databases. The second tier performs real-time clinical event detection based on real-time vital sign data collected from on-body wireless sensors attached to those high-risk patients. We employ machine-learning techniques to analyze data from both tiers, assigning scores to patients in real time. The assigned scores can then be used to trigger early-intervention alerts. Preliminary study of an early warning system component and a wireless clinical monitoring system component demonstrate the feasibility of this two-tiered approach. 

... Read complete abstract on page 2.
Toward a Two-Tier Clinical Warning System for Hospitalized Patients

Complete Abstract:

Clinical study has found early detection and intervention to be essential for preventing clinical deterioration in patients at general hospital units. In this paper, we envision a two-tiered early warning system designed to identify the signs of clinical deterioration and provide early warning of serious clinical events. The first tier of the system automatically identifies patients at risk of clinical deterioration from existing electronic medical record databases. The second tier performs real-time clinical event detection based on real-time vital sign data collected from on-body wireless sensors attached to those high-risk patients. We employ machine-learning techniques to analyze data from both tiers, assigning scores to patients in real time. The assigned scores can then be used to trigger early-intervention alerts. Preliminary study of an early warning system component and a wireless clinical monitoring system component demonstrate the feasibility of this two-tiered approach.
Abstract

Clinical study has found early detection and intervention to be essential for preventing clinical deterioration in patients at general hospital units. In this paper, we envision a two-tiered early warning system designed to identify the signs of clinical deterioration and provide early warning of serious clinical events. The first tier of the system automatically identifies patients at risk of clinical deterioration from existing electronic medical record databases. The second tier performs real-time clinical event detection based on real-time vital sign data collected from on-body wireless sensors attached to those high-risk patients. We employ machine-learning techniques to analyze data from both tiers, assigning scores to patients in real time. The assigned scores can then be used to trigger early-intervention alerts. Preliminary study of an early warning system component and a wireless clinical monitoring system component demonstrate the feasibility of this two-tiered approach.

Introduction

Within the medical community, there has been significant research into preventing clinical deterioration among hospital patients. Clinical study has found that 4–17% of patients will undergo cardiopulmonary or respiratory arrest while in the hospital[1]. Early detection and intervention are essential to preventing these serious, often life-threatening events. Indeed, early detection and treatment of patients with sepsis has already shown promising results, resulting in significantly lower mortality rates[2].

Two distinct approaches have emerged to support these early detection efforts. First, existing electronic medical records may be processed using machine learning techniques to identify early warning signs of clinical deterioration[3]. Second, real-time data collection systems based on wireless sensor network (WSN) technologies have been proposed to collect vital sign data at a high granularity[4,5]; in turn, this data may be used as inputs to automated scoring systems[6].

The application of either approach in isolation has significant limitations. Early warning systems based on existing medical records suffer from the sparseness of measurements; in general hospital units, patients’ vital signs are typically collected manually by a nurse, at a granularity of only a handful of readings per day. While real-time monitoring systems fill the gap in measurements, it is nevertheless impractical to intensively monitor all patients in general hospital units. Such an approach would introduce issues of scalability and complexity; it would also inconvenience many patients, who would be asked to wear sensors even if they were not necessary.

In this paper, we envision a novel two-tier system for clinical early warning and intervention. Our proposed system combines the two approaches in order to achieve the benefits of both (identifying clinical events) while avoiding their individual shortcomings. At the first tier, advanced machine learning techniques are used to identify at-risk patients from real-time data in existing electronic medical records. At the second tier, these at-risk patients are provided WSN-based monitoring devices which collect high-fidelity vital sign data in real time. These tiers are joined by a predictive algorithm which scores a patient’s predicted outcome based on both real-time vital sign data and the coarser-grained data found in electronic medical records.

System Design

In contrast to earlier research that focuses on wireless networking or machine learning algorithms alone, we propose an integrated system to provide clinical decision support to health care professionals. Our system consists of the following key components:
1. **Early Warning System (EWS):** The EWS identifies at-risk patients from existing real-time Electronic Medical Record (EMR) data using machine learning algorithms.

2. **Wireless Clinical Monitoring System (WCMS):** The at-risk patients identified by the EWS are provided wireless sensors to collect vital sign data at high sampling rates. The vital sign data are then streamed to EMR database systems using a wireless sensor network.

3. **Real-Time Event Detection System (RES):** An event detection algorithm running on EMR database servers analyzes the electronic medical records and the streaming vital sign data in real time, in order to identify clinical deterioration events. When an event is identified, the RES automatically notifies nurses through the hospital’s paging system to provide early intervention.

In the remainder of this paper, we will focus on the design and feasibility study of the first two components. The Real-Time Event Detection System component remains an important part of future work.

**Early Warning System**

The first step in providing early intervention lies in determining which patients are most at risk of clinical deterioration and who would most benefit from more intensive monitoring. In order to make this decision, our proposed system will leverage existing data from electronic medical record systems. As hospitals transition to electronic record-keeping, their databases are aggregating an increasing wealth of information about a patient’s history and current condition: demographic data, the results of lab tests, information about prescriptions, low-granularity vital sign data, etc. The challenge lies in determining how all the factors represented in this data may serve as potential warnings of clinical deterioration. We envision the use of advanced machine-learning techniques to find correlations between medical data and the patient’s future outcome. These correlations can then be used to identify at-risk patients based on the data contained in their medical records.

In order to demonstrate the feasibility of an Early Warning System component, we applied logistic regression\[7\] techniques to predict patients’ outcomes (specifically, whether or not they would be transferred to the ICU) on a historical dataset. This dataset cataloged 28,927 hospital visits from 19,116 distinct patients between July 2007 and January 2010. It contained a wide array of demographic and medical data for each of the visits, such as age, manually-collected vital sign data, pharmacy data, and whether or not the patient was transferred to the ICU. All data contained in this dataset were taken from historical EMR databases, and reflects the kinds of data that would realistically be available at the first tier of our proposed clinical warning system. This study serves as a proof-of-concept for our vision of using machine learning to identify at-risk patients and (ultimately) to perform real-time event detection.

We will begin this section by providing a description of the logistic regression approach used in this study. As part of this description, we will describe several important pre-processing steps needed to adapt these techniques for our application. We will then describe the performance of this approach under two different scenarios: a retrospective analysis where each patient is assigned a single score, and a real-time simulator where patient scores are continually recomputed as new EMR data is entered.

**Algorithm Overview.** At a high level, logistic regression predicts an outcome $y$ from a vector of input variables $x_1, x_2, \ldots, x_k$. Each element in the vector $x$ represents one reading of a particular kind; for example, $x_1$ may indicate a patient’s blood oxygen saturation while $x_2$ may be the patient’s temperature. The predicted outcome $y \in (0, 1)$ indicates the likelihood of some event occurring based on the input variables. $y$ values near 1 represent a strong prediction of the event occurring (e.g., the patient being admitted to the ICU), while values near 0 represent a high likelihood of the event not occurring. The variable $y$ is discretized using a simple thresholding procedure. The threshold controls the trade-off between false positive and false negative rates (i.e., higher thresholds decrease false positives but increase false negatives).

In order to compute predictions, a set of weights $\beta_1, \beta_2, \ldots, \beta_k$ are used to determine the importance of each parameter that is part of vector $x$. Large positive $\beta$ values indicate that a particular variable is highly predictive of an outcome of interest, while large negative $\beta$ values indicate that a variable is highly protective against this outcome. For example,
when \( x_1 \) represents a patient’s blood oxygen saturation, a large negative \( \beta_1 \) would indicate that patients with higher blood oxygenation levels are less likely to be ultimately admitted to the ICU.

The logistic equation is formally defined as follows. For each case (hospital visit), we have an input vector \( x_1, x_2, \ldots, x_k \). We also have a set of coefficients \( \beta_1, \beta_2, \ldots, \beta_k \) and an intercept value \( \beta_0 \) which are constant across all cases. From these inputs, we may compute

\[
z = \beta_0 + \sum_{i=1}^{k} \beta_i x_i
\]

for each case. The output \( z \) is in the range \((-\infty, \infty)\). We then confine the prediction to the range \((0, 1)\) by computing

\[
y = \frac{1}{1 + e^{-z}}
\]

The vector \( \beta \) is produced offline from a training (historical) dataset. Each case in the training dataset has both an input vector \( x \) and a known outcome \( y \). Since the outcome is discrete and known, \( y \) may only take the values of 1 (positive outcome) or \(-1\) (negative outcome). \( \beta \) may then be fitted to the training dataset using standard techniques, such as least-squares.

While developing our study, we identified several additional, domain-specific preprocessing measures which enhanced the predictive performance of the basic logistic regression algorithm. These steps are described below.

**Time division.** By itself, logistic regression does not operate on time-series data. That is, each variable input to the logistic equation corresponds to exactly one data point: e.g., a blood pressure variable would consist of a single blood pressure reading. In a clinical application, however, it is important to capture unusual changes in vital sign data over time. Such changes may precede clinical deterioration by hours\(^8\), providing a chance to intervene if detected early enough. In addition, not all readings in time-series data should be treated equally; the value of some kinds of data may change depending on its age. For example, a patient’s condition may be better reflected by a blood-oxygenation reading collected one hour ago than a reading collected 12 hours ago.

To capture the temporal effects in our data, we retain a sliding window of all the collected data points within the last 24 hours. We then subdivide this data into a series of \( n \) equally-sized buckets (e.g., 6 sequential buckets of 4 hours each). In order to capture variations within a bucket, we compute three values for each bucket: the minimum, maximum, and mean data points\(^1\). Each of the resulting \( 3n \) values are input to the logistic regression equation as separate variables.

**Handling missing data.** A complication of using clinical data is that not all patients will have values for all variables. Many types of clinical data involve lab tests that are not routinely performed on all patients. This problem is compounded by dividing time into buckets: even when a patient has had a particular lab test, it will only provide a data point for one bucket.

To deal with missing data points, we first “carry over” historical values. Specifically, when a bucket is empty, we insert the patient’s most recent reading from any time earlier in the dataset, possibly going all the way back to the beginning of the hospital stay. To handle cases where a patient had no previous data for a particular variable, we also calculate the mean value for each variable over the entire historical dataset. These mean values are used as a fallback when the carrying-over procedure cannot find any data to carry over.

**Performance results: retrospective analysis.** After implementing the above algorithm in MATLAB, we evaluated its accuracy over the historical dataset. Specifically, logistic regression was used to predict whether the patient was admitted to the ICU.

For the purposes of training, we looked at a single 24-hour window of data from each patient. For patients admitted to ICU, this window consisted of 26 hours – 2 hours prior to ICU admission; for all other patients, this window consisted of the first 24 hours’ of their hospital stay. We subdivided this 24-hour window into 6 contiguous windows of 4 hours each.

\(^1\)We attempted to add a fourth value for standard deviation, but found that our historical data set was too sparse to calculate \( \sigma \) for nearly all buckets. We are considering other ways to measure variability as future work.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respirations (bucket 6 max)</td>
<td>4.45</td>
</tr>
<tr>
<td>Oxygen Saturation, pulse oximetry (bucket 6 min)</td>
<td>-4.22</td>
</tr>
<tr>
<td>Shock Index (bucket 6 max)</td>
<td>4.01</td>
</tr>
<tr>
<td>Respirations (bucket 6 mean)</td>
<td>3.41</td>
</tr>
<tr>
<td>BP, Systolic (bucket 6 min)</td>
<td>-2.96</td>
</tr>
<tr>
<td>Coagulation modifiers (bucket 1)</td>
<td>2.70</td>
</tr>
<tr>
<td>Pulse (bucket 6 max)</td>
<td>2.55</td>
</tr>
<tr>
<td>Respirations (bucket 6 min)</td>
<td>2.51</td>
</tr>
<tr>
<td>BP, Diastolic (bucket 6 max)</td>
<td>2.48</td>
</tr>
<tr>
<td>Oxygen Saturation, pulse oximetry (bucket 6 mean)</td>
<td>-2.44</td>
</tr>
</tbody>
</table>

Table 1: The 10 highest-weighted variables from the training dataset

As noted above, the logistic model requires training data to fit the unknown $\beta$ values from known data and outcomes. We divided the dataset into two halves; the first half was used to train the model. Table 1 provides a sample of the output from the training process, listing the 10 highest-weighted variables and their coefficients. We then used the second half of the dataset as testing data. We generated a predicted outcome for each case in the testing data, using the $\beta$ weights derived from the training data. As noted above, we then applied various thresholds to convert these predictions into binary values, and compared the results against the ground-truth ICU outcome.

Figure 1: ROC curve of logistic model’s predictive performance under retrospective analysis.

Figure 1 plots the ROC curve of the model’s performance under a range of thresholds. The X axis plots 1 - the specificity for a given threshold. The Y axis plots the corresponding sensitivity at the same threshold. Both performance metrics are important from a clinical perspective. Higher X values correspond to a larger number of false alarms; if alarms are too frequent, clinicians may be more inclined to ignore them. Higher Y values correspond to more patients correctly being identified for early intervention.

Based on these results, a threshold of $y = 0.9323$ was chosen to achieve a specificity close to 95%. This specificity value was chosen in turn to generate only a manageable number of alerts per hospital floor per day. Even at this relatively high specificity, the logistic regression approach achieves a sensitivity of 48.8%. Table 2 summarizes the
Table 2: The predictive performance of logistic regression at a chosen cutpoint under retrospective analysis

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area under curve</td>
<td>0.8834</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.9500</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.4877</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>0.3138</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>0.9753</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.9292</td>
</tr>
</tbody>
</table>

Performance at this cutpoint under several common statistical metrics.

*Performance results: real-time simulation.* In order to train the logistic model, our retrospective analysis looked at only a single 24-hour window of data for each patient. In a system that predicts patients’ outcomes in real time, their scores would be recomputed each time new data is entered into the EMR database. Hence, each patient will effectively have a series of scores over the length of their hospital stay, and an alarm will be triggered when *any one* of these scores is above the threshold.

To better understand the model’s performance under these circumstances, we performed a second evaluation using a real-time simulator. The simulator “replays” each patient’s EMR data in chronological order over the length of their entire hospital stay. For each entry in the EMR dataset, the simulator collects all the patient’s data over a 24-hour window ending at the new entry, and computes a new score for the patient (using the $\beta$ values fitted during the retrospective study). ICU admission is predicted if any one of the patient’s scores exceed a given threshold.

Figure 2: ROC curve of logistic model’s predictive performance under real-time simulation.

For this real-time simulation, we replayed a smaller, more recent data set cataloging 1,284 hospital visits from 1,204 distinct patients between October 2010 and December 2010. Figure 2 plots the corresponding ROC curve of the model’s performance.

As with the retrospective study, we chose a threshold to achieve a specificity close to 95%, yielding a cutpoint of $y = 0.9987$. At a specificity of 94.9%, the model achieves a sensitivity of 41.3%. Table 3 summarizes the model’s performance at this cutpoint.
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Area under curve</td>
<td>0.7293</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.9492</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.4127</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>0.2955</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>0.9691</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.9229</td>
</tr>
</tbody>
</table>

Table 3: The predictive performance of logistic regression at a chosen cutpoint under real-time simulation

**Wireless Clinical Monitoring System**

Up-to-date vital sign data plays an important role in identifying early-warning signs of clinical deterioration. In Intensive Care Units (ICUs), high-fidelity wired equipment is employed to collect vital sign data in real time. However, in general hospital units, vital sign data is typically collected manually by a nurse. The labor involved means that only a handful of readings may be collected per day.

Recent work has explored the deployment of real-time patient monitoring systems based on WSN technology\[^9,10,11,4,5\]. WSNs consist of small, embedded devices equipped with wireless radios and moderate onboard processing capabilities. A salient feature of WSNs for clinical monitoring applications is that they can be deployed with minimal fixed infrastructure. Communication is performed entirely over a low-power wireless radio, and nodes may operate from an onboard battery. Hence, a WSN-based monitoring system can be deployed in general hospital floors, where rooms may lack dedicated communication wiring. Moreover, such devices may be used by patients without restricting their mobility; this is particularly useful in general hospital units, where mobility can be an important part of some patients’ recovery process.

Figure 3 pictures a WSN-based wireless pulse oximeter device developed by Washington University around the TelosB\[^12\] WSN platform. This device has characteristics and capabilities typical of many WSN-class devices. Data collection is driven by an embedded MSP430 microcontroller, which is interfaced with a Smith Medical OEM pulse-oximeter sensor using a custom circuit board. The device periodically collects pulse and blood oxygenation level data, and routes the data to a wired access point using an onboard radio based on the IEEE 802.15.4 radio standard. WSN network coverage is achieved by plugging additional TelosB nodes into electrical outlets in patients’ rooms and in the hallway. These new nodes will autonomously locate other nearby nodes and participate in routing sensor data from patient nodes to a wired access point, where they are entered into the EMR database system.

The lack of wired infrastructure in these systems poses significant engineering challenges. WSN platforms typically feature small, fixed energy supplies (often AA or 9V batteries) which place severe constraints on energy consumption. In addition, the low-power radio links formed by these devices often exhibit poor reliability and/or strongly temporal properties. To deal with these challenges in a patient monitoring system, we have previously developed a routing strategy based around the *Dynamic Relay Association Protocol* (DRAP)^[13]. Under this strategy, a dedicated
Routing infrastructure is formed by plugging relay nodes into available electrical outlets. The Collection Tree Protocol (CTP)\[14\] is employed on these relay nodes to establish a data collection tree rooted at a wired base station. Patient nodes use DRAP to locate nearby relay nodes. When patient nodes are ready to transmit vital sign data, they transmit directly to one of these nearby relay nodes; the relay nodes then route the data to a wired base station. An important benefit of this approach is that it greatly simplifies power management on patient nodes. Because patient nodes do not actively participate in routing other nodes’ traffic, they only need to power on their radios to discover relay nodes or to transmit their own sensor data. Moreover, because the relay nodes are plugged into electrical outlets, they may remain fully active at all times to route the patient nodes’ traffic.

In previous work, we have carried out a clinical trial of a real-time patient monitoring system in a step-down unit at Barnes-Jewish Hospital\[5\]. For completeness, we will briefly summarize the results of this study here.

![Figure 4: Deployment of the wireless clinical monitoring system in the step-down unit of Barnes-Jewish Hospital.](image)

46 patients were enrolled in the study between June 4, 2009 and January 31, 2010. The step-down unit where the study took place has 16 patient rooms and covers an area of 1200 m\(^2\). We deployed 18 relay nodes to provide coverage within the unit, as shown in Figure 4 (the blue square denotes the base station while red circles denote relay nodes). During the trial, the custodial staff unplugged the relay nodes occasionally to power their cleaning equipment. In addition, two relays were destroyed by impact with cleaning equipment. Due to the redundancy of the deployed relays, neither of these events had adverse effects on network reliability.

**System reliability.** The combination of CTP and DRAP achieved high network reliability, with a median of 99.68% of packets successfully delivered to the base station. Network outages were infrequent and had a 95%-percentile time-to-recovery of 2.4 minutes. Indeed, the reliability of the system was dominated by sensor reliability: a median of 80.85% packets contained valid sensor readings, with the others flagged as invalid by the pulse oximeter. Retroactive analysis suggested oversampling and sensor disconnection alarms could be used to increase sensor reliability\[13\].

**System lifetime.** Patient nodes achieved a lifetime of up to 69 hours from a 9V battery power source.

**Related Work**

Numerous projects in existing literature propose using inexpensive wireless hardware to assist in healthcare. Most closely related to our proposed system are CodeBlue\[9\], Alarm-NET\[10\], MEDiSN\[4\], and our own recent clinical monitoring system\[5\], which deploy WSN-based devices for real-time patient monitoring. A common theme among these systems is that they focus primarily on systems issues related to collecting raw sensor data in a wireless sensor
network: hardware design, network protocols, and system reliability. In this paper, we envision a complete system for clinical event detection; real-time vital sign data would be mined along with existing electronic medical records to predict patient outcome. A real-time collection scheme forms an important component of our proposed system, and would leverage the contributions made by these prior studies.

The development of algorithms for detecting clinical deterioration has a long tradition in medical literature. Two types of algorithms for detecting clinical deterioration may be distinguished: algorithms that rely on medical knowledge and general machine learning techniques that do not require domain knowledge. A number of scoring systems that use medical knowledge exist for various medical conditions. For example, the effectiveness of Severe Community-Acquire Pneumonia (SCAP) and Pneumonia Severity Index (PSI) in predicting outcomes in patients with pneumonia is evaluated in[15]. Similarly, outcomes in patients with renal failure may be predicted using the Acute Physiology Score (12 physiologic variables), Chronic Health Score (organ dysfunction), and age (APACHE) score[16]. However, these algorithms are best suited for specialized hospital units. In contrast, the detection of clinical deterioration on general hospital units requires more general algorithms. For example, the Modified Early Warning Score (MEWS)[6] uses systolic blood pressure, pulse rate, temperature, respiratory rate, age, and BMI to predict clinical deterioration. These physiological and demographic parameters may be collected at bedside, making MEWS suitable for general hospital units.

An alternative to algorithms that use medical knowledge is to adopt standard machine learning techniques. This approach has two important advantages over traditional rule-based algorithms. First, it allows us to consider a larger number of parameters during the prediction of patient outcomes. Second, since they do not use a small set of rules for predicting outcomes, it is possible for machine learning approaches to achieve improve accuracy. Learning techniques such as decision trees[3], neural networks[17], and logistic regression[18] have been used to identify clinical deterioration. A distinguishing feature of our approach is the use of a two-tier detection mechanism for identifying clinical deterioration, which leverages both existing EMR data and real-time sensor data.

Conclusion

In this paper, we discussed our vision for a novel two-tier system clinical early warning and intervention. We proposed an integrated architecture for detecting clinical events based on a combination of electronic medical records and real-time vital sign data. Our proposed system consists of three key components: (1) an Early Warning System which identifies at-risk patients from existing electronic medical records; (2) a Wireless Clinical Monitoring System which collects real-time vital sign data from those at-risk patients; and (3) a Real-Time Event Detection System which analyses both electronic medical records and real-time vital sign data to identify clinical deterioration events. A feasibility study of our Early Warning System shows promising results. A retrospective study over a historical dataset of 28,927 hospital visits predicted patient ICU transfer with a sensitivity of 48.8% and a specificity of 95%, while real-time simulation over a recent dataset of 1,284 hospital visits predicted patient ICU transfer with a sensitivity of 41.3% and a specificity of 95%. This work represents a promising step toward early clinical warning that has the potential to significantly improve the quality and outcome of patient care in hospitals.

This publication was made possible by Grant Number UL1 RR024992 from the National Center for Research Resources (NCRR), part of the National Institutes of Health (NIH) and NIH Roadmap for Medical Research. Its contents are solely the responsibility of the authors. Additional funding was provided by the Barnes-Jewish Hospital Foundation and NSF through NeTS-NOSS Grant CNS-0627126 and CRI Grant CNS-0708460.

References


