Identification of Temporal Changes on Patients at Risk of LONS with TPRMine: A Case Study in NICU

Catherine Inibhunu, Member IEEE
Ontario Tech University
Oshawa, Ontario, Canada

Carolyn McGregor AM, Senior Member IEEE
Ontario Tech University
Oshawa, Ontario, Canada
University of Technology Sydney
Ultimo, NSW, Australia

***Abstract*—A neonatal intensive care unit (NICU) provides specialized care for preterm or ill term infants. The onset of many conditions they can develop are not obvious to physicians until they are significantly impacted and this could result in death. An example of such a problem is neonatal infection which is a common cause of death for premature infants. It remains a challenging task for clinicians to accurately diagnose the presence of bacteria on patients with frequent presence of multiple comorbidities. There is potential for early detection of neonatal infections by timely analysis of patient physiological data and this can lead to improved health outcome of critically ill patients. This paper demonstrates application of a method for Temporal Pattern Recognition and Mining (TPRMine) in order to (a) understand if continuous analysis of temporal changes in patient physiological data streams can lead to discovery of pathophysiological patterns from patients at risk of neonatal sepsis and, (b) utilize the resulting analysis for formulating and testing hypothesis facilitating statistical quantification of patients.**

Keywords—Temporal Pattern Mining, Temporal Analysis, Patient Scoring, Pattern Recognition in Healthcare

I. INTRODUCTION

Neonatal sepsis is a bacterial infection in infants admitted to NICUs and can be a source of severe conditions or lead to a loss of life. Premature infants are extremely susceptible to infectious pathogens and diagnosis of conditions like neonatal sepsis is a challenging problem as the signs of the condition are often not specific. The infections could occur in the mother's uterus, during birth or at postnatal. Very low birth weight (VLBW) infants especially those born significantly premature and weighing less than 1500 grams have a higher risk of neonatal infection, as incidence of infection in preterm infants is 3 to 10 times higher, compared to full term infants, as noted in Griffin & Moorman [1] . Of the VLBW infants, 25% can be infected with late-onset neonatal sepsis (LONS) after 3 days of life and this is a key contributor to the doubling of incidence of death as well as 50% increase in hospital stay for these infants [2]. Diagnosis of neonatal sepsis is normally difficult to establish. Griffin and Moorman (2001) hypothesized that detection of abnormal characteristics of the heart rate (HR) could precede clinical diagnosis of neonatal sepsis. The researchers noted that if such information is gathered early, it can be useful in addition to standard clinical parameters for care of patients [1]. Joshi, et al (2019) notes that analysis of heart rate variability (HRV) from preterm infants can aid in tracking changes in subclinical signatures of the disease [3].

LONS, which occurs in approximately 10% of neonates and 25% of VLBW infants hospitalized in NICUs is the focus of this paper. McGregor, Catley & James (2012) noted that slight changes in patients physiological data, which may not be obvious to detect through traditional manual recordings at regular intervals, could be essential in discovering the onset of sepsis in neonates [4]. These researches developed an algorithm for detecting LONS by measuring the variability of the HR and the respiratory rate (RR). In prior work McGregor et al [4] indicated that there was a significant increase in the probability of LONS when there was a classification of less than 38% in HRV and the respiratory rate variability (RRV) did not fall below that same level. In that work, patient data analyzed was sampled at 2 readings a minute due to limitations of available data. As a result, patterns that could be derived from a higher sampling rate were not possible.

In NICUs, high frequency data are collected from monitors and sensors that provide measures relating to patients' vital status as described in [5]. Timely analysis of this data has the potential to provide valuable insights, which can be crucial when making critical decisions for the care of premature and ill term infants. In particular, every second, a single patient in an NICU has physiological data generated from multiple sensors at the bedside and displayed on Phillips Intellivue monitors. The same data is simultaneously captured in Artemis cloud using the real time data streaming modules described in Inibhunu, et al., [5]. This data is comprised of multiple physiological features such as HR, RR, blood oxygen saturation (SpO_2), pulse rate from Plethysmogram (PulsePleth) along with other features. Each of these 4 data streams have a tuple generated approximately every second (1024 ms) and therefore these are termed as high frequency time series data streams. Within 1 hr, approx. 3518 independent tuples are generated for each of these data streams i.e. HR, RR, SpO_2 and PulsePleth resulting in a total of 14,400 records per patient. If there are 51 NICU beds in a hospital and all beds occupied by neonatal patients, then a total of 792,000 records are captured per hour for HR, RR, SpO_2 and PulsePleth.

Research still remains open for development of methods that can detect changes and relationships in patterns that may be exhibited in high frequency patient data. The ability to detect hidden relationships or patterns from such data could lead to discovery of unknown diseases or condition pathophysiology.

We have proposed a knowledge discovery method by augmenting several machine learning principles creating the TPR process and an associated TPRMine algorithm as described in [6]. In this paper, we demonstrate the application of TPRMine to assess whether continuous analysis of temporal changes in patient's HR, RR, SpO_2 and PulsePleth represent pathophysiological patterns in those physiological data streams for patients at risk of LONS.

The rest of the paper is organized as follows: section II describes the TPRMine method and its application for enhancement of the threshold method in [4]. The results are provided in section III, discussion is in section IV and conclusion and future works in section V.

II. METHODS

TPRMine is a method for knowledge discovery that adopts a stepwise approach to temporal pattern discovery from time oriented data by first applying a scaled mathematical formulation of the time series data. This process was accomplished by modelling the problem space as a finite state machine representation where for a given timeframe, a time series data segment can transition from one state to the next based on some associated weights as described in [7].

In this paper, we demonstrate the many states a patient may transition to in a given period of time, thereby highlighting important information on patients that may be at risk of LONS. This is accomplished in this research by utilizing physiological data streams captured from the Artemis Big Data Platform in [5] under ethics approved in (HiREB 3859-D) and Ontario Tech (REB #14136) in addition to Late Neonatal Sepsis study under (HiREB 4833-C) and Ontario Tech (#15536).

As described in [6], TPRMine algorithm contains stepwise modules each with its own analytical process. In this paper, we demonstrate application of TPRMine using data captured from NICU patients. To facilitate the process, enhancements to the threshold scoring method in [4] together with a more granular data representation demonstrate the application of TPRMine algorithm to multiple physiological data streams from patients in NICU. The goal is to demonstrate how this approach could enable discovery of temporal patterns within the hour of high or low threshold.

A. Background on HRV and RRV Algorithms

The HRV and RRV algorithms detailed in [4] calculate variability of a patient's heart and respiratory rates and generate a percentage threshold as shown in Figure 1. A patient is then assigned a threshold score based on the percentage of HR and RR variability. These algorithms were demonstrated as able to detect potential LONS on those patients with a HRV < 38% and RRV > 38%. These are the patients who would fall in the Score D square in Figure 1.

Figure 1: Threshold Scoring Algorithms

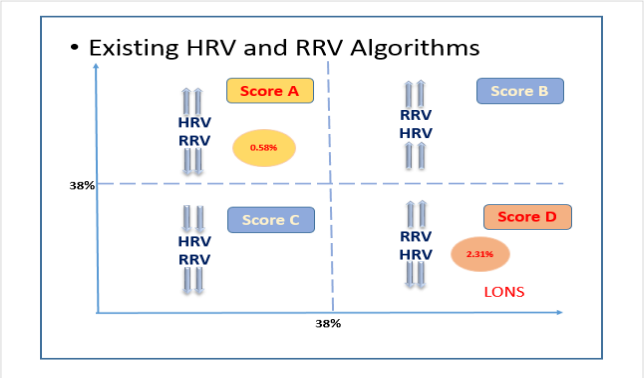
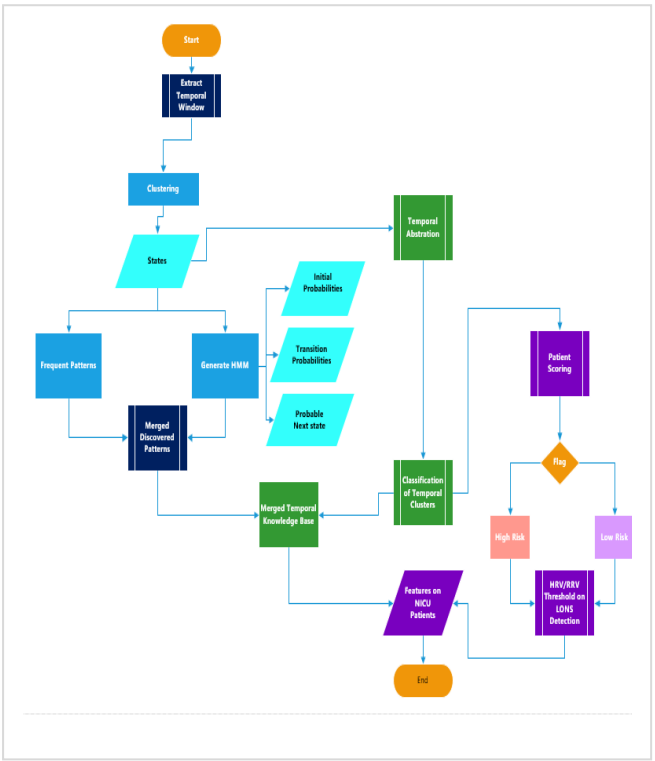


Figure 2 presents the stepwise application of TPRMine to multiple patients' data streams augmented with the threshold scoring algorithms presented in Figure 1. First, TPRMine formulates temporal data windows that are processed in sequences and outputs from one module are fed as input to the subsequent analytical modules until all data is processed. The result is a knowledge base that includes data driven clusters that forms the states a patient may transition to in a given time period, the predicted next state and then augmenting these with clinical context quantifying each state with a vital score.

Figure 2: TPRMine Process Applied to NICU Data
Purple indicates augmenting the TPRMine modules with HRV and RRV Algorithms



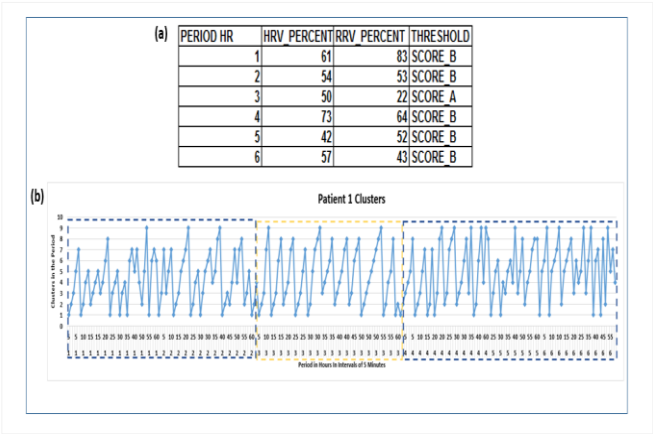
III. RESULTS

An 11 hour subset of the dataset was used for this experiment. During those 11 hours, a total of 5,702,400 million records were generated from 36 patients who were in the NICU. The results from processing all and subsets of this data with TPRMine is presented as follows.

A. Clustering with TPRMine Algorithm

First is the identification of the states that patients transition to in a given period and incorporating the HRV and RRV threshold scores. Figure 3 shows results from 6 hours of a random patient where (a) shows the threshold scores based on the HRV, RRV algorithms where the patient has been in two scoring regions A and B and (b) shows the results from TPRMine on the different states transitioned. Figure 4 shows similar results for a patient with a threshold Score D throughout the 6 hour period.

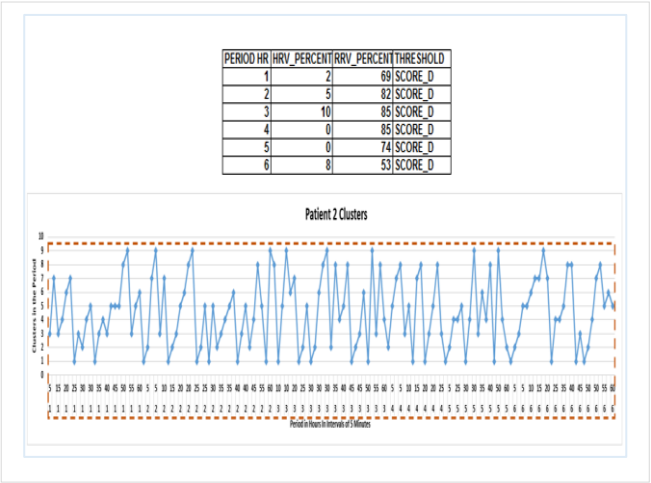
Figure 3: Patient with Scores A/B, State Transitions in 6 hours



B. Patient Scoring with TPRMine Algorithm

Next results shows each patient scored using the TPRMine algorithm where by derived states are quantified with clinical context thus identifying risky or non-risky states as described in [6]. Results are presented in Figure 5 where (a) identifies several facts about a specific period of time; the cluster, the cluster centroids and the vital score of the particular cluster, and (b) determines what is the percentage of times a patient stayed at a specific vital score. For example in the PeriodHr 2, the patient's HR, RR or SpO_2 were abnormal the entire hour as the patient had 0% of vital score 0, 92.68% in vital score 1 and 7.32% in vital score 2.

Figure 4: Patient with Score D and State Transitions in 6 hours



C. Comparison of two Patients State Transitions with TPRMine

TPRMine identifies the many states the patient transitions to within a selected time interval. A further comparison on another set of two random patients state transitions using 6 hours of data are presented in Figure 6, where (a) shows the number of states the patient transitions in each 5 minute temporal window. Figure 6 (b) presents the state transitions between the 2 patients on 6 hours of data. Patient 102 (identified as having threshold scores A or B) shows higher number of state transitions in each period hour compared to patient 101 (identified as having threshold scored D).

Figure 5: Generating Patient Vital Scores with Clinical Context

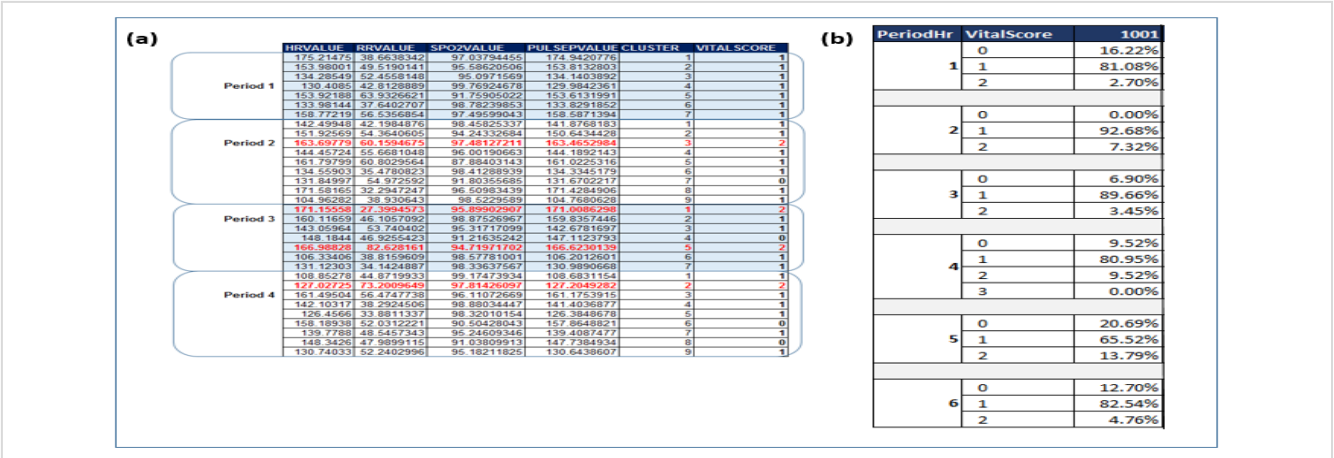
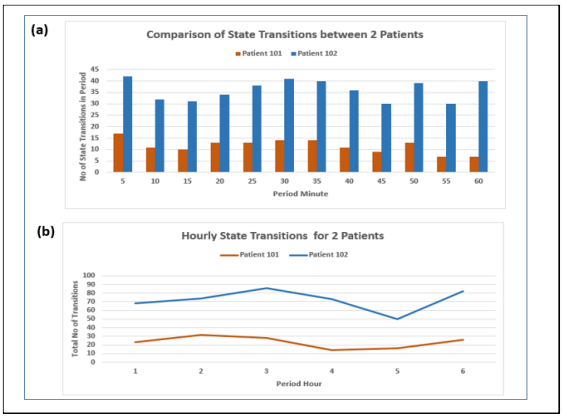


Figure 6: Number of State Transitions in a 6 hour Period between 2 Randomized Patients



D. Comparison of the Entire Patient Population and Formulation of Hypothesis Tests

After comparing the two different patients, an analysis of the entire patient population is performed to understand the variation on the number of state transitions across all the patient cohort over an 11 hour period. As there were 36 patients in the data analyzed, 3 were found to have a threshold score D while 32 had a threshold score B and 1 had a threshold score A. Two groups were then formed, those with threshold score D in group B while the rest were placed in group A. Differences were noted on the results from analysis of the two patient groups. In particular, using 6 hours of data statistical principles for group analysis was completed by calculating hourly average transitions and standard deviation between the two patient groups. The results showed that hourly average of state transitions in group A is consistently higher than 50 compared to that of group B in each hour analyzed.

Using a randomized sample, a null hypothesis test was formulated to test whether the average number of state transitions is the same between the 2 patient groups. The results indicated a difference on the means between the two patient groups with a p-value = 0.003288.

IV. CONCLUSION AND FUTURE WORK

This paper demonstrates application of a temporal pattern recognition and mining algorithm for analysis of patients physiological data thus enhancing the threshold scoring approach in McGregor et al [4]. This process is completed by integrating a more granular data representation to demonstrate that application of TPRMine to multiple physiological data streams captured from patients in NICU could enable discovery of temporal patterns within the hour of high or low threshold.

This is accomplished by generation of data driven clusters including artefacts such as cluster means and cluster variances which can be utilized for retrospective analysis. Additionally, the data driven clusters form the states that a patient in NICU may transition to in a given period. Quantification of the total number of states a patient may transition to in a given period of time is completed using a scoring mechanism.

These results indicates a difference in the average state transition between the two groups of patients and could facilitate formulation of null hypothesis tests. Such a process can enable clinical case studies that could lead to discovery of pathophysiological patterns in physiological data streams captured from patients at risk of conditions such as LONS.

During our research, we have discovered that there are fundamental differences among patients admitted for care. In particular, there are neonates who are preterm and those that are full term infants, preterm have VLBW and are 3 to 10 times more likely to be predisposed to neonatal sepsis as compared to full term infants with normal birthweight [8]. To this respect the temporal behaviors on data generated from such cohort of patients can be very different.

As such, TPRMine can therefore be applied separately to these different groups of patients resulting in differing knowledge base. In the future we look to utilize TPRMine to understand temporal patterns on varying cohort of neonates at risk of LONS.

V. REFERENCES

[1] M. Griffin and J. Moorman , "Toward the early diagnosis of neonatal sepsis and sepsis-like illness using novel heart rate analysis," *Pediatrics,* pp. 97-104, 2001.

[2] P. Griffin, M. O'Shea , A. Bisonette , F. Harrell , G. Lake and J. Moorman , "Abnormal Heart Rate Characteristics PrecedingNeonatal Sepsis and Sepsis-Like Illness," *Pediatric Research ,* p. 920–926 , 2003.

[3] R. Joshi, K. Deedee, C. Guo, J. Bikker, L. Feijs, C. van Pul and P. Andriessen, "statistical Modeling of Heart Rate Variability to Unravel the Factors Affecting Autonomic Regulation in Preterm Infants," *Scientific Reports,* pp. 1-9, 2019.

[4] C. McGregor, C. Catley and A. James, "Variability Analysis with Analytics Applied to Physiological Data Streams," in *25th International Symposium on Computer-Based Medical Systems (CBMS)*, Rome, Italy, 2012.

[5] C. Inibhunu, R. Jalali, I. Doyle, A. Gates, J. Madill and C. McGregor, "Adaptive API for Real-Time Streaming Analytics as a Service," in *EMBC*, Berlin, 2019.

[6] C. Inibhunu, "A Temporal Pattern Discovery Method and its Application to Physiological Data: A Case Study," *Doctorial dissertation,* 2020.

[7] C. Inibhunu and McGregor, "State Based Hidden Markov Models for Temporal Pattern Discovery in Critical Care," in *IEEE Life Sciences Conference (LSC)* , Montreal, Canada, 2018.

[8] A. v. d. Hoogen, Infections in neonatal intensive care, Prevalence, Prevention and Antibiotic use, ISBN: 978-90-393-5119-2, 2009.