High level descriptions generated from time oriented raw patient data collected from various medical systems is called temporal abstractions. Medical systems generate data every millisecond, therefore the volume of abstractions grows exponentially. Clinicians could benefit having all this information about patients in aid of decision making; however, a huge challenge faces anyone trying to understand these abstractions.

In this paper we look at how temporal data mining as a knowledge retrieval mechanism on time oriented data can be applied on generation and interpretation of abstractions created from raw patients data thereby presenting meaningful information to clinicians.

I. INTRODUCTION

In the evolution of health informatics, and the introduction of electronic medical records, monitoring systems and clinical decisions support systems, patient data is being collected everywhere and anyone trying to make sense of this data is faced with huge challenges. This is evident more so for health care provider who need to make critical decisions about patients [1, 2].

In neonatal intensive care units (NICU) where monitors and sensors generate data every millisecond, temporal abstraction has attempted to bridge the gap between raw data to deriving abstracts that could be more understandable. However, given the huge volumes of data, intelligent systems are needed to handle this kind of data in aid of knowledge extraction that can aid in critical decision making that can help improve the care of patients.

Temporal data mining, a more recent field of knowledge extraction, has a main objective of mining large sequential data while maintaining the temporal nature of the data. This is quite different from statistical time-series analysis in that in temporal data mining the ordering of data and relationships of events is crucial in the modelling process [3].

II. Temporal Data Mining

Artemis [2] is an example of a framework built in clinical setting for acquisition of streams of real time sensor data, there is a need for creating processes that can handle data from different systems and with differing levels of granularity thereby generating high level abstractions that could be meaningful to healthcare providers in aid of decision making. In the generation of temporal abstractions, a challenge exists on how to deal with the huge volumes of abstractions generated. Temporal data mining which incorporates the temporal nature of the data aids in knowledge extraction and discovery from data and has been identified in various works as a credible process for mining time oriented data [4].

III. Temporal Data Mining Process

Mining of temporal data takes on several activities; data selection, pre-processing, transformation, data mining, interpretation and visualization.

The selection process ensures the right amount of data is identified, whether this is during variable selection from structured databases or temporal abstractions from time-oriented from real time or static systems. Picking the appropriate data is crucial in the output generated from any mining algorithm. In the pre-processing phase a lot of data cleaning could happen depending on the domain of the data been mined.

During temporal abstraction where the right intervals and windowing mechanisms is key for generating meaningful abstracts, the pre-processing phase is quite critical. The outcome of any mining algorithm is depended a lot on what is taken in the first stages during mining process. Transforming data can be done to put it in a format ready for the data mining process. An example can be the creation of abstracts which can then be input for the data mining step.

Interpreting outputs from a temporal mining framework is not trivial, especially where streams of data are captured in real time and the need to maintain the temporal nature of the data. Interpretation of knowledge extracted is another challenge. Depending on the clinical domain where mining is performed, there is need for incorporation of a clinical knowledge base. Medical ontologies can serve as a source of a medical knowledge base and included in data validation of outputs from the data mining process.

III. Why Temporal Data Mining

Temporal data mining is quite different from regular data mining in that
maintaining the order in which data is captured is crucial for algorithms that try to identify association and patterns from the data over a period of time. In clinical settings, where raw data is at the lowest granularity, there are several challenges on how one can extract hidden knowledge from this data. Temporal abstractions systems generate high level granularity of data with an aim of providing meaningful data. However the creation of these abstracts is quite a challenge; data can be of various formats and structures, and could be captured in very different systems that are either real time or static.

There are several aspects that have been identified in the literature as key factors for any temporal data mining framework; scalability to real time data streams, dealing with data anomalies, domain knowledge base and semantics, visualization of the extracted data, and how to handle data of different data types, differing systems distributed or not.

During the pre-processing phase where appropriate sequencing of time oriented data is critical for generating high level granularities of data for feeding to mining algorithms. Understanding the trends on temporal data is another aspect that can be aided by data mining where different levels of abstractions can be done to form higher order levels of granularity [6, 7].

Healthcare organizations faced with increasing volumes of patients’ data need to have frameworks that have mining algorithms capable of scaling and thereby be able to handle static or real time data streams captured in distributed or standalone systems. Pattern recognition is also a key factor in temporal data mining where identifying different trends in data could result to bringing out unknown knowledge. In critical care, pattern recognition can help in detection of unknown conditions and therefore call for appropriate treatment modalities.

IV: Techniques in Clinical Temporal Data Mining

Given the huge amount of data processed in clinical care settings, it’s quite critical that the right mining technique is implemented. The adoption of the association rules and the frequent pattern mining is seen in various frameworks, others includes Bayesian networks where temporal nodes represent relationships on events and change of state in temporal data has been used. A review on Bayesian networks is done in [7]. Probabilistic models are still a dominant part of data mining and have been used in various clustering and classifications algorithms. Clinical data analysis of temporal abstracts is also seen as part of knowledge retrieval A review of systems that utilize this approach was reported in [5].

IV. CONCLUSION

Anyone trying to make sense of the huge volumes of medical data about patients generated every millisecond from sensors and bedside monitors is faced with several computational challenges from receipt of the data to its interpretations. Temporal abstracts have attempted to bridge the gap between raw to higher granular data except that, generating abstracts from time oriented data is not a trivial exercise. Temporal data mining as a knowledge retrieval mechanism can be incorporated in various stages of abstractions creation. However, several challenges face processes trying to extract knowledge from temporal data such as: the ability to process ever increasing data volumes; maintaining the temporal nature of the data; the ability to scale to evolving systems, data and processes; filtering and pruning; the identification of appropriate patterns; incorporating medical domain knowledge; and the interpretation of generated results.

Our review shows temporal data mining can be used to generate and interpret temporal abstraction derived from raw patient data. Extending mining approaches to be able to handle real-time data streams from multi-agent medical systems is a promising research area.

Acknowledgment
This research is sponsored by Dr Carolyn McGregor Canada Research Chair in Health Informatics at University of Ontario Institute of Technology

V. REFERENCES


